Etude des processus rédactionnels d'un texte : les données temporelles comme indices de la gestion des unités de production

Emilie Ailhaud

▶ **To cite this version:**

Emilie Ailhaud. Etude des processus rédactionnels d'un texte : les données temporelles comme indices de la gestion des unités de production. Linguistique. Université de Lyon, 2016. Français. NNT : 2016LYSE2161 . tel-01485691

HAL Id: tel-01485691
https://theses.hal.science/tel-01485691

Submitted on 9 Mar 2017

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.

N° d'ordre NNT : 2016LYSE2161

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 484 Lettres, Langues, Linguistique et Arts

Discipline : Sciences du langage

Soutenue publiquement le 12 décembre 2016, par :

Emilie AILHAUD

Étude des processus rédactionnels d'un texte : les données temporelles comme indices de la gestion des unités de production

Devant le jury composé de :

Claire DOQUET, Professeure des universités, Université Paris 3, Présidente

Thierry OLIVE, Chargé de recherche HDR, C.N.R.S., Rapporteur

Florence CHENU, Ingénieur de recherche, C.N.R.S., Examinatrice

Jacques DAVID, Maitre de conférences, Université de Cergy-Pontoise, Examinateur

Harriet JISA, Professeure émérite, Université Lumière Lyon 2, Directrice

Denis ALAMARGOT, Professeur des universités, Université Paris 12, Co-directeur

Doctorat de l'Université de Lyon

Délivré par l'Université Lumière Lyon 2
École doctorale Lettres, Langues, Linguistique & Arts (3LA)
Laboratoire Dynamique du Langage (UMR 5596)

Étude des processus rédactionnels d'un texte : les données temporelles comme indices de la gestion des unités de production

---

Thèse de doctorat en Sciences du Langage

Par Emilie AILHAUD

Présentée et soutenue publiquement le 12 décembre 2016

Composition du jury :

Denis ALAMARGOT, Professeur des universités, ESPE Académie de Créteil, UPEC	Co-directeur
Florence CHENU, Ingénieur de recherche, CNRS – Laboratoire DDL, Lyon	Examinatrice
Jacques DAVID, Maitre de conférences, ESPE Académie de Versailles, UCP	Examinateur
Claire DOQUET, Professeur des universités, Université de la Sorbonne Nouvelle	Rapporteure
Harriet JISA, Professeur émérite, Université Lyon 2	Directrice
Thierry OLIVE, Chargé de recherche (Habilité), CNRS – Laboratoire CeRCA, Poitiers	Rapporteur

Résumé

Étude des processus rédactionnels d'un texte : les données temporelles comme indices de la gestion des unités de production

Ce travail de thèse s'appuie sur les données temporelles issues de l'enregistrement de deux corpus, afin de mieux comprendre la gestion des processus impliqués dans la production de textes. Le premier corpus était constitué de textes rédigés par des enfants de 10 à 15 ans, alors que les textes du second corpus ont été produits par des jeunes adultes. Nous avons utilisé les données chronométriques – durées des pauses et vitesse d'écriture – en les mettant en lien avec le contenu linguistique du texte – liage syntaxique et sémantique, choix lexicaux. Nous avons également analysé les révisions portées sur le texte, en examinant l'ordre de ces opérations, leur durée et le lieu où la révision était initiée. Le but de notre travail était de mieux saisir quelles étaient les unités de production et comment les processus rédactionnels tels que la micro-planification, la formulation et la révision étaient gérés au sein de ces unités. D'un point de vue développemental, nous avons mis en évidence le passage d'une planification portant sur la clause à une planification portant sur une unité plus large ; ce développement a ensuite été mis en lien avec la maitrise des normes typographiques de segmentation des textes. Les données des adultes ont révélé une anticipation de la planification des clauses lorsqu'elles montraient un lien syntaxique fort avec les clauses environnantes ; en outre, il a été montré que la production était rendue plus facile ou plus couteuse en fonction de la présence de certaines caractéristiques linguistiques. Enfin, l'analyse des révisions a permis d'affiner certains résultats, et a en outre révélé la présence probable de relectures, majoritairement effectuées lors des pauses entre les phrases, dont le rôle serait sans doute d'aider à la planification du texte.

Mots - clés : production écrite, texte, développement, pauses, révisions, cohésion, unités, planification

The study of processes in written text production: chronometric data as clues to understanding units of production

The work presented in this thesis is based on chronometric data examined in two corpora with the goal of better understanding the processes involved in written text production. The first corpus is composed of texts written by children and adolescents (10 to 15 years of age), while the texts used in the second corpus were written by university students. Chronometric measures – pause length and writing speed – were examined and related to the linguistic structures – lexicon, syntactic and semantic connectivity – observed in the texts. Revisions of the texts were also examined taking into account their location, their length, and the order in which they were produced. The goal of the work undertaken was to better understand the units of production and how writing processes such as micro-planning, formulation and revisions are managed within these units. From a developmental point of view our analysis reveals an increase in the size of the unit, moving from planning clause by clause to planning over multiclausal structures. A relationship between this development and increasing control over conventions of punctuation to segment texts was also revealed. In the adult data we show the anticipation of planning over multiclausal structures displaying syntactic connectivity. In addition, we show how certain characteristics of linguistic structures can facilitate or hinder production. Finally, the analysis of revisions permits a refining of previous results and suggests rereading during pauses between sentences, which probably helps in text planning.

Key-words: written production, text, development, pauses, revision, cohesion, units, planning

Remerciements

Mes remerciements vont tout d'abord à Harriet Jisa, qui a dirigé cette thèse. C'est à vous que je dois mes débuts dans la recherche : depuis mon arrivée en première année de master vous m'avez fait découvrir avec passion les thématiques telles que le développement tardif du langage ou l'écriture, et vous m'avez formée au travail de recherche. Merci de m'avoir fait confiance, en m'ayant permis de faire cette thèse sous votre direction, et en m'ayant poussée à rencontrer des chercheurs en France et en Europe. Merci surtout pour votre bienveillance et votre disponibilité, qui ont permis que je prenne tant de plaisir à faire ce travail.

Merci à Denis Alamargot, qui a accepté de co-diriger cette thèse. Les échanges que nous avons eus m'ont permis d'avancer dans mes réflexions, et j'espère que nous continuerons à avoir des discussions fructueuses.

Merci à Florence Chenu, présente également depuis le début de mes recherches. Tu as su répondre avec justesse à mes multiples questions, et me pousser à toujours m'en poser davantage. Merci d'avoir été aussi disponible pour m'accompagner dans ce travail.

Merci à Claire Doquet, Thierry Olive et Jacques David, qui ont accepté de faire partie de mon jury. Je vous remercie d'avoir pris le temps de lire et examiner ce travail. Votre expertise permettra certainement de soulever de nombreuses discussions, et j'espère que nous aurons encore l'occasion d'avoir des échanges riches autour de ce sujet qui nous réunit avec passion : l'écriture.

Merci aux personnes ayant contribué aux collectes des données utilisées dans ce travail, notamment dans le cadre des projets ReFlex et DyTO, ainsi qu'à tous les participants ayant accepté d'y prendre part.

Cette thèse s'est déroulée dans les meilleures conditions possibles grâce au financement accordé par le Labex ASLAN, ainsi qu'à l'accueil au sein du laboratoire Dynamique du Langage. Je remercie donc particulièrement François Pellegrino et Sophie Kern. Votre manière de concevoir la recherche comme un travail exigeant et audacieux, l'attention portée à la valorisation scientifique, et surtout l'humanité avec laquelle vous dirigez ces structures ont profondément marqué ma vision du travail de chercheur.

Au sein du DDL, je voudrais tout d'abord remercier les personnes ayant permis que, d'un point de vue « matériel », cette thèse se déroule au mieux : notamment, merci à Linda et Rabia pour toute la gestion administrative ; merci à Christian pour ton soutien en informatique et ton soutien moral grâce aux pauses-cafés de début de journée… Je ne pourrais pas faire la liste exhaustive de la manière dont chaque membre du DDL a, d'une manière ou d'une autre, contribué à ce travail. Je peux cependant évoquer les enseignants-chercheurs qui, lors de mon master puis lors de discussions au cours de mon doctorat ont fait croitre mes connaissances et mon intérêt pour la linguistique. Merci également aux *membres de l'axe Dendy*, ainsi qu'aux personnes ayant animé les ateliers méthodologiques, pour les échanges riches et les formations dispensées.

Ce travail n'aurait pas pu voir le jour sans l'accompagnement des personnes ayant partagé avec moi ces années d'études, que ce soit en licence à Nice (Loan, Cyrielle) ou en master à Lyon (Lucie, Marion, Ophélie, Jeanne, Anne-Éléonore, Néo, Ivan) : nous avons partagé les plaisirs liés à la découverte des sciences du langage, les difficultés à l'approche des examens, et tant de moments joyeux qui m'ont permis de passer des belles années à l'université : merci à vous !

L'aventure s'est poursuivie au DDL avec les doctorant(e)s du labo. Merci à Natacha, Sara et Natalia, qui m'ont accueillie dans leur bureau alors que je débutais cette thèse : grâce à vous j'ai pu tout de suite me sentir « chez moi » au labo, et vos discussions sur les descriptions de langue me permettaient de m'évader un peu de ma psycholinguistique ! Merci aux autres doctorants du labo, avec qui j'ai notamment partagé de nombreuses pauses déjeuners et discussions, que ce soient les « anciennes » doctorantes qui étaient là à mon arrivée, ou celles et ceux qui sont arrivés depuis (Rozenn, Bénédicte, Noellie, Anne-Laure, Marion, Yoon-Mi, Darine, Roula, Khadija, Léa, Noé, Jinke, Denis…) : c'était une chance de vous rencontrer, et je suis toujours émerveillée de nos différences, qui reflètent si bien la richesse et la diversité de notre laboratoire. Enfin, un merci particulier à l'équipe des doctorantes de Dendy : merci à Jennifer, pour ta disponibilité et l'attention que tu portes à chacun au labo ; merci à Camille, pour nos discussions passionnées en psycholinguistique et tout ce que nous avons partagé depuis le master ; merci à Hélène et à Agathe, pour les pauses-thés, les rires, et nos discussions de filles qui permettent de se changer les idées ; merci à Ludivine et à Leslie, pour votre présence et votre amitié qui permettent de partager les moments de joie comme les moments de doute dans notre beau bureau ensoleillé ; merci à Marie, pour tes mille petits mots d'encouragement et ta disponibilité, notamment durant cette dernière ligne droite partagée ensemble… merci les p'tits chats !

Cette thèse n'aurait pas été menée à bien sans le soutien indéfectible des amis qui m'ont accompagnée durant ces années lyonnaises : merci à Olivier, pour ton accueil et ton attention permanente ; merci à Aurore, Clément, Marie, Fabiola, Anne-Claire, Claire-Marie, Pierre, Eloïse, Antoine, Claire, Hugues … vous tous qui avez été présents durant ces années, et auprès de qui j'ai toujours trouvé une oreille attentive et pu me ressourcer. Sans oublier Flore, dont l'amitié m'est si précieuse depuis de très nombreuses années.

Enfin, cette thèse n'aurait sans doute pas commencé sans l'accompagnement de ma famille. Merci à mes oncles et tantes, à mon parrain et à ma marraine, vous qui depuis plusieurs années me soutenez, en étant certes parfois étonnés de mes choix professionnels, mais en m'encourageant pour mener à bien ce travail. Merci à Grand-Maman, pour tes attentions, ton écoute, ta bienveillance permanente. Merci à Maxime et Aurélien, pour votre intérêt empreint de curiosité envers mon travail, et tous les petits témoignages de soutien apportés pendant ces longues années. Merci à Adeline pour ta présence réconfortante, nos nombreuses discussions et confidences, et particulièrement pour ton aide lors de la fin de cette rédaction de thèse. Et surtout, merci à Papa et Maman. Vous m'avez toujours fait confiance dans mes choix d'étude, même quand vous pensiez que d'autres auraient pu être meilleurs. C'est grâce à cette confiance, ainsi qu'à l'amour que vous m'avez toujours apporté, que j'ai pu mener mon petit chemin, explorer ce qui me passionnait, et que je peux aujourd'hui présenter cette thèse… merci pour tout cela.

C'est bien sûr à Thibaut que j'adresse mes derniers remerciements. Tu es arrivé « en cours de route » de cette thèse, découvrant ainsi ce qui pouvait se cacher derrière les Sciences du langage. Merci pour l'intérêt que tu as porté à mon travail, lorsque j'essayais maladroitement de t'expliquer en quoi il consistait. Merci pour ta patience et tout ce que tu as fait ces dernières semaines, lors de cette « dernière ligne droite » qui a semblé bien longue. Merci surtout pour ta présence quotidienne et l'amour que tu m'apportes depuis maintenant deux ans : c'est grâce à toi que j'ai cet équilibre, grâce auquel j'ai pu aller au bout de ce travail. Merci…

Table des matières

Résumé 3
Remerciements 4
Table des matières 6
Table des illustrations - Figures 9
Table des illustrations - Tableaux 12
Introduction 15
Cadre théorique 21

Chapitre 1 - Produire un texte : la complexité du produit fini 22
1. D'une juxtaposition de clauses à un texte cohérent 23
1.1. De la clause au texte 23
1.2. L'assemblage des clauses 26
1.3. La constitution d'un schéma textuel 29
2. Le contenu linguistique du texte 31
2.1. Éléments linguistiques susceptibles d'être analysés 31
2.2. Différences dues au type de texte 40
2.3. Différences dues à l'âge 41
3. Les unités de segmentation du texte 45
3.1. Les unités identifiées lors de l'étude d'un texte 45
3.2. Les difficultés de segmentation lors de l'écriture d'un texte 46
4. Conclusion 50

Chapitre 2 - Produire un texte : la complexité des processus rédactionnels 51
1. Le modèle de Hayes et Flower et son évolution 52
2. L'identification des processus 56
2.1. « Planification » et « formulation » 56
2.2. « Révision » 61
3. La gestion des processus 65
3.1. Les contraintes de capacité cognitive 65
3.2. Interactions entre les processus rédactionnels 66
3.3. Gestion temporelle 69
4. Différences interindividuelles 72
4.1. Développement de l'expertise 72
4.2. Profils de scripteurs 74
5. Conclusion 76

Chapitre 3 - Étude du décours temporel de l'écriture 77
1. Méthodes d'analyse de la gestion temporelle des processus rédactionnels 78
1.1. Divers protocoles d'étude 78
1.2. Les retours sur le texte 81
1.3. Conclusion sur les méthodes à notre disposition 84
2. L'identification des unités de production grâce aux données chronométriques 86
2.1. La question des unités de production 86

2.2. Lien entre données chronométriques et contenu linguistique 87
3. Vers de nouvelles questions de recherche 97
3.1. Questions émergeant des études présentées 97
3.2. Un défi méthodologique : l'analyse des données chronométriques 98
3.3. Présentation de notre recherche 101
Méthodologie générale 105

Chapitre 4 – Méthodologie générale 106
1. Recueil et traitement des données 107
1.1. Recueil des données du corpus « ReFlex » 107
1.2. Recueil des données du corpus « Dimitri » 109
1.3. Extraction des données 112
2. Codage général 114
2.1. Données linguistiques : les grandes unités de segmentation et leurs relations 114
2.2. Données chronométriques 122
3. Analyses utilisées 126
3.1. Explication de la variance de données continues 126
3.2. Établissement de profils 129
Résultats 133

Chapitre 5 – Les unités de production : aspect développemental 134
1. Une perspective développementale des unités de production 136
1.1. Introduction 136
1.2. Méthodologie 137
1.3. Résultats 140
1.4. Discussion 146
2. La phrase : de la maitrise de ses conventions graphiques à sa pertinence cognitive 150
2.1. Introduction 150
2.2. Méthodologie 151
2.3. Résultats 154
2.4. Discussion 158
2.5. Conclusion 160
3. Discussion générale 162
3.1. Vers une planification au-delà de la clause 162
3.2. Perspectives de recherche et limites du corpus 163

Chapitre 6 – Données chronométriques du corpus adulte 165
1. Description générale des données chronométriques 166
1.1. Ponctuation 167
1.2. Gestion des pauses et des temps d'exécution 170
1.3. Évolution de la dynamique d'écriture au long du texte 178
1.4. Résumé des profils 183
1.5. Discussion sur la description générale des données chronométriques 191
2. Les unités de segmentation du texte 197
2.1. Longueur des pauses en fonction de l'emplacement syntaxique 197
2.2. Longueur des pauses en fonction du lien syntaxique entre les clauses 202

3. Conclusion 214

Chapitre 7 - Lien entre rythme d'écriture et contenu linguistique des unités 215
1. Méthodologie 216
1.1. Codage du contenu linguistique 216
1.2. Préparation des analyses 241
2. Temporalité de l'écriture des phrases et des clauses 245
2.1. Hypothèses 245
2.2. Écriture de la phrase 246
2.3. Écriture des clauses 253
2.4. Discussion générale 266
3. Conclusion 271

Chapitre 8 – Étude des révisions 272
1. Révisions du corpus 273
1.1. Codage des révisions 273
1.2. Hypothèses 276
1.3. Description générale 276
1.4. Localisation des révisions 278
1.5. Gestion temporelle des révisions 281
1.6. Discussion 283
2. Étude de cas 288
2.1. Présentation du texte étudié 288
2.2. Le produit final 289
2.3. Le manuscrit 292
2.4. Étude du « film de l'écriture » 295
2.5. Étude des pauses 302
2.6. Discussion 307
3. Discussion générale 315
Conclusion 317
1. Synthèse des résultats 318
1.1. Les unités de production : de la clause à la phrase 318
1.2. La gestion des unités de production 320
1.3. Les données temporelles comme indices des processus rédactionnels 322
2. Perspectives 324
Références 327
Annexes 349

Table des illustrations - Figures

Figure 1. Schématisation des composantes d'une séquence narrative d'après Adam (2011) 36
Figure 2. Modèle de Hayes et Flower (1980) 52
Figure 3. Modèle de Hayes (1996) 54
Figure 4. Modèle de Hayes (2012) 55
Figure 5. Modèle de Levelt (1999) 57
Figure 6. Modèle en cascade de Olive (1994) 68
Figure 7. Consignes données pour la production des textes - corpus « ReFlex » 108
Figure 8. Stimulus utilisé pour la tâche « Le Petit Chaperon Rouge » 110
Figure 9. Texte d'amorce proposé pour la tâche « Dimitri » 110
Figure 10. Exemple de détourage des séquences d'un texte avec le logiciel Eye & Pen 113
Figure 11. Exemple de profils établis par la classification hiérarchique 131
Figure 12. Données chronométriques en fonction de l'ordre de production (moyennes en ms) 141
Figure 13. Données chronométriques en fonction du type de texte (moyennes en ms) 141
Figure 14. Données chronométriques en fonction du niveau scolaire (moyennes en ms) 141
Figure 15. Durée des pauses initiales en fonction de l'ordre de production et du niveau scolaire (moyennes en ms) 142
Figure 16. Durée des pauses initiales en fonction des emplacements syntaxiques «I», «C» et «T» et du niveau scolaire (moyennes en ms) 143
Figure 17. Durée des pauses initiales en fonction des emplacements syntaxiques «T-P» et «T+P» et du niveau scolaire (moyennes en ms) 143
Figure 18. Durée des pauses initiales des clauses indépendantes et dépendantes (moyennes en ms) 144
Figure 19. Durée des pauses initiales des clauses dépendantes matrices et non-matrices (moyennes en ms) 145
Figure 20. Durée des pauses initiales des clauses dépendantes en fonction du type de construction syntaxique (moyennes en ms) 145
Figure 21. Profils liés aux problèmes de ponctuation 155
Figure 22. Profils liés à la longueur des pauses suivant l'emplacement syntaxique 157
Figure 23. Répartition de la ponctuation en fonction de l'emplacement syntaxique (en %) 168
Figure 24. Distribution du temps de pause avant et après les signes de ponctuation en fonction de l'emplacement syntaxique (moyennes en %) 168
Figure 25. Profils liés à la ponctuation (moyennes en %) 169
Figure 26. Durée de la pause initiale en fonction de l'emplacement syntaxique (moyennes en ms) 171
Figure 27. Profils liés à la moyenne des pauses et du temps d'exécution 172
Figure 28. Proportion de pauses par catégorie de durées (durées en ms, moyennes en %) 173
Figure 29. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique (durées en ms, moyennes en %) 174
Figure 30. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique – Profil 1 (durées en ms, moyennes en %) 175
Figure 31. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique – Profil 2 (durées en ms, moyennes en %) 175
Figure 32. Proportion de pauses par grande catégorie de durée en fonction de l'emplacement syntaxique (durées en ms, moyennes en %) 176
Figure 33. Répartition du temps d'écriture (moyennes en %) 177
Figure 34. Profils liés à la répartition du temps d'écriture (moyennes en %) 178
Figure 35. Profils liés à la proportion de texte écrit dans chaque partie (moyennes en %) 179

Figure 36. Différences entre la moyenne des pauses de chaque partie et la moyenne générale des pauses du texte (moyennes en ms) 180
Figure 37. Profils liés à la différence entre la moyenne des pauses de chaque partie et la moyenne générale des pauses du texte (moyennes en ms) 181
Figure 38. Profils liés à la proportion d'écriture consacrée à l'exécution par partie (moyennes en %) 182
Figure 39. Proportion de pauses de chaque emplacement syntaxique par quartile (en %)..... 198
Figure 40. Proportion de pauses de chaque quartile par emplacement syntaxique (en %)..... 198
Figure 41. Proportion de pauses de chaque emplacement syntaxique par décile (en %) 199
Figure 42. Durée des pauses en fonction de l'emplacement syntaxique (moyennes en ms) 200
Figure 43. Durée de la pause initiale de la clause en fonction du numéro de la clause dans la phrase (moyennes en ms)..... 208
Figure 44. Durée de la pause initiale de la clause en fonction de son aspect matrice ou non matrice (moyennes en ms)..... 208
Figure 45. Durée de la pause initiale de la clause en fonction de son aspect obligatoire ou facultatif (moyennes en ms)..... 208
Figure 46. Durée de la pause initiale de la clause en fonction du type syntaxique (moyennes en ms) 208
Figure 47. Durée moyenne de l'écriture interne de la clause en fonction de sa profondeur syntaxique (en ms/seq) 209
Figure 48. Durée moyenne de l'écriture interne de la clause en fonction du numéro de l'unité A dans la phrase (en ms/seq) 209
Figure 49. Durée moyenne des pauses inter-mots de la clause en fonction de sa profondeur syntaxique (en ms/seq) 210
Figure 50. Durée moyenne des pauses inter-mots de la clause en fonction du numéro de l'unité A dans la phrase (en ms/seq) 210
Figure 51. Durée moyenne des pauses intra-mots de la clause en fonction de sa profondeur syntaxique (en ms/car) 211
Figure 52. Durée moyenne des pauses intra-mots de la clause en fonction du numéro de l'unité A dans la phrase (en ms/car)..... 211
Figure 53. Durée moyenne du temps d'exécution de la clause en fonction du type de subordonnée (en ms/car)211
Figure 54. Temporalité de l'action 233
Figure 55. Distribution des scripteurs en fonction du nombre de révisions dans le texte 277
Figure 56. Répartition des révisions en fonction des types d'opérations (en %)..... 277
Figure 57. Répartition des révisions en fonction du type d'éléments modifiés (en %) 277
Figure 58. Type d'opération de révision en fonction de l'élément révisé 277
Figure 59. Distribution des catégories de révision en fonction de la distance entre l'origine et le site (en %).... 279
Figure 60. Distribution la distance entre l'origine et le site en fonction des catégories de révision (en %) 279
Figure 61. Localisation des sites et origines des révisions en fonction des parties du texte 280
Figure 62. Localisation des catégories de révisions en fonction des parties du texte..... 280
Figure 63. Répartition des pauses en fonction des catégories de révision (durée en quartile) 283
Figure 64. Répartition des pauses en fonction des opérations de révision (durée en quartile)..... 283
Figure 65. Répartition des pauses en fonction de la distance entre l'origine et le site de la révision (durée en quartile) 283
Figure 66. Texte du scripteur S31 - Produit final..... 289
Figure 67. Texte du scripteur S31 - Manuscrit..... 292
Figure 68. Nombre de révisions du texte du scripteur S31 en fonction du type d'opération et de segment révisé 297
Figure 69. Distribution des révisions du texte du scripteur S31 et du corpus en fonction du type d'opération (en %) 297

Figure 70. Distribution des révisions du texte du scripteur S31 et du corpus en fonction du type de segment révisé (en %) 297
Figure 71. Distance entre l'origine et le site de la révision - texte du scripteur S31 et corpus (répartition en %)298
Figure 72. Répartition des pauses initiales en fonction de l'emplacement syntaxique (durées en quartiles et déciles) 302
Figure 73. Contenu des pauses D10 (en %) 303
Figure 74. Reprises esthétiques initiées durant l'écriture du texte du scripteur S31 311
Figure 75. Capture d'écran du logiciel Eye & Pen - Aperçu des mouvements oculaires recueillis lors de la production d'un texte 312
Figure 76. Origine et site des révisions du texte du scripteur S31 312

Table des illustrations - Tableaux

Tableau 1. Terminologie liée aux processus rédactionnels 60
Tableau 2. Ordre de passage lors du recueil des données « ReFlex » 108
Tableau 3. Segmentation en clauses – Résumé des critères 117
Tableau 4. Codage syntaxique des clauses – Résumé 122
Tableau 5. Données chronométriques en fonction de l'ordre de production, du type de texte et du niveau scolaire (moyennes en ms, écart-types et effectifs) 141
Tableau 6. Durée des pauses initiales en fonction de l'emplacement syntaxique et du niveau scolaire (moyennes en ms, écart-types et effectifs) 143
Tableau 7. Durée des pauses initiales des clauses en fonction du type de clause (moyennes en ms, écart-types et effectifs) 144
Tableau 8. Durée des pauses initiales des clauses matrices et non-matrices (moyennes en ms, écart-types et effectifs) 144
Tableau 9. Segmentation des textes en fonction du niveau scolaire (moyennes et écart-types) 154
Tableau 10. Seuil des 20% des pauses les plus longues des textes en fonction du niveau scolaire (moyennes et écart-types) 154
Tableau 11. Proportion de problèmes liés à la ponctuation en fonction du niveau scolaire (moyennes en % et écart-types) 155
Tableau 12. Répartition des scripteurs dans les profils liés aux problèmes de ponctuation en fonction du niveau scolaire 156
Tableau 13. Proportions de la longueur des pauses suivant l'emplacement syntaxique en fonction du niveau scolaire (moyennes en % et écart-types) 156
Tableau 14. Répartition des scripteurs dans les profils liés à la longueur des pauses suivant l'emplacement syntaxique en fonction du niveau scolaire 157
Tableau 15. Croisement des profils liés à la ponctuation et à la longueur des pauses 158
Tableau 16. Nombre de signes de ponctuation et répartition du temps de pause en fonction de l'emplacement syntaxique (moyennes et écart-types) 168
Tableau 17. Ponctuation - Caractéristiques des profils (V1) 170
Tableau 18. Durées des pauses et des temps d'exécution (moyennes en ms et écart-types) 170
Tableau 19. Moyenne des pauses et du temps d'exécution par profil (en ms) 172
Tableau 20. Moyenne des temps de pause et d'exécution - Caractéristiques des profils (V2) 172
Tableau 21. Proportion de pauses par catégorie de durée (en ms) en fonction de l'emplacement syntaxique (moyennes en % et écart-types) 173
Tableau 22. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique – Moyennes par profils (durée en ms, moyennes en %) 174
Tableau 23. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique - Caractéristiques des profils (V3) (durée en ms, moyennes en %) 176
Tableau 24. Répartition générale du temps d'écriture (moyennes en % et écart-types) 177
Tableau 25. Répartition du temps d'écriture – Caractéristiques des profils (V4) 178
Tableau 26. Longueur des parties (moyennes et écart-types) 178
Tableau 27. Longueur des parties - Caractéristique des profils (V5) 179
Tableau 28 Différences entre la moyenne des pauses de chaque partie et la moyenne générale des pauses du texte (moyennes en ms et écart-types) 180
Tableau 29. Durée des pauses par partie - Caractéristiques des profils (V6) 181
Tableau 30. Répartition du temps d'écriture par partie (moyennes en % et écart-types) 181
Tableau 31. Proportion d'écriture consacrée à l'exécution par partie - Caractéristiques des profils (V7) 182

Tableau 32. Résumé des variables chronométriques 184
Tableau 33. Liens entre les variables 186
Tableau 34. Lien entre les profils : moyenne des temps de pauses et exécution 187
Tableau 35. Lien entre les profils : comportement suivant les parties 189
Tableau 36. Répartition des pauses du texte en fonction de leur emplacement syntaxique et de la durée en quartile 197
Tableau 37. Répartition des pauses en frontière syntaxique (« P », « A », « C ») les plus longues (D8 à D10). 199
Tableau 38. Pause initiale de la clause - Effets fixes du modèle 2 207
Tableau 39. Facteurs liés au lien entre les clauses expliquant le mieux la variance des données chronométriques des clauses 213
Tableau 40. Codage de la profondeur syntaxique des syntagmes - Exemple 217
Tableau 41. Mots grammaticaux et lexicaux 218
Tableau 42. Natures grammaticales relevées pour les syntagmes du corpus 220
Tableau 43. Rôle sémantique des compléments circonstanciels 225
Tableau 44. Classement sémantique des mots de liaison 226
Tableau 45. Récapitulatif du codage des syntagmes 231
Tableau 46. Critères d'assignation de la temporalité de l'action 233
Tableau 47. Catégories sémantiques des prédicats - Exemples 235
Tableau 48. Unités étudiées concernant la continuité des sujets 236
Tableau 49. Continuité des sujets - Exemple 236
Tableau 50. Compléments de lieux "externes" et "corporels" - Exemples 236
Tableau 51. Champs lexicaux des référents - Exemples 238
Tableau 52. Analyses des phrases et des clauses - Groupes créés 242
Tableau 53. Variables chronométriques (VD) liées aux phrases 246
Tableau 54. Variables linguistiques (VI) liées aux phrases 247
Tableau 55. Récapitulatif des effets des VI sur les VD liées aux phrases 249
Tableau 56. Sens des effets des facteurs linguistiques sur les données chronométriques liées à la phrase 251
Tableau 57. Variables chronométriques (VD) liées aux clauses 253
Tableau 58. Variables linguistiques (VI) liées aux clauses 254
Tableau 59. Récapitulatif des effets des VI sur les VD liées aux clauses matrices 257
Tableau 60 Récapitulatif des effets des VI sur les VD liées aux clauses subordonnées 258
Tableau 61. Récapitulatif des effets des VI sur les VD liées aux clauses autonomes 260
Tableau 62. Variables linguistiques - effets observés sur les clauses 263
Tableau 63. Distance entre l'origine et le site de la révision en fonction des catégories de révision 278
Tableau 64. Localisation des révisions en fonction des parties du texte 279
Tableau 65. Répartition des pauses avant l'origine et la révision en fonction du type d'opération, de la distance entre l'origine et le site et de la durée de la pause (en quartile) 282
Tableau 66. Révisions identifiées dans le manuscrit 293
Tableau 67. Révisions en début du texte - Détails de l'étape 1 299
Tableau 68. Seuil des quartiles et déciles - Texte du scripteur S31 et corpus (en ms) 302
Tableau 69. Révisions en début du texte - Détails de l'étape 1 avec durée des pauses (en ms) 304
Tableau 70. Révisions en début du texte - Détails de l'étape 2 avec durée des pauses (en ms) 304
Tableau 71. Révisions en début du texte - Détails de l'étape 3 avec durée des pauses (en ms) 306
Tableau 72. Révisions autour de la mention "petit Dimitri" - Détails avec durée des pauses (en ms) 306

Introduction

L'activité de production écrite est complexe et l'amélioration des compétences rédactionnelles se poursuit tout au long de la vie : de l'écriture des lettres de son prénom en maternelle à la production de textes dans le cadre professionnel, le scripteur doit apprendre à adapter son activité pour parvenir à transmettre le message souhaité, tout en ayant des ressources cognitives limitées. Les contraintes pesant sur le scripteur lors de la production écrite d'un texte sont en effet multiples (Plane, Olive & Alamargot, 2010). Il doit tout d'abord gérer des contraintes sur le plan linguistique : chaque système linguistique a ses propres caractéristiques syntaxiques, lexicales et orthographiques à respecter ; par ailleurs, un texte n'étant pas une simple suite de mots ou de phrases, le scripteur est contraint par le texte déjà écrit puisqu'il doit assurer un lien sémantique et syntaxique entre ses phrases ; enfin, il est contraint par les prescriptions données par la consigne ou qu'il se fixe, afin de construire le type de texte attendu et adapté au destinataire. De plus, le scripteur doit gérer des contraintes sur le plan cognitif, puisque les capacités de son système cognitif sont limitées (McCutchen, 1996) et qu'il doit gérer de nombreux processus rédactionnels dont certains demandent davantage de ressources, notamment lorsqu'ils ne sont pas tout à fait automatisés. D'après le modèle de Hayes et Flower proposé en1980, le scripteur doit effectuer la préparation conceptuelle de ses propos, établir des choix lexicaux et syntaxiques afin de transposer ces idées en une séquence linguistique, puis relire son texte pour réaliser les corrections nécessaires.

Dès lors, au vu des contraintes pesant sur le scripteur, on peut se demander quelles sont les unités de production, et comment cette production est gérée. Autrement dit, sur quelle étendue de texte le scripteur effectue-t-il ses différents choix, avant de mettre par écrit ce qu'il a préparé, et pourquoi certains segments sont-ils plus difficiles à écrire que d'autres ? La question des « unités » de production est au cœur de préoccupations linguistiques et psychologiques. En effet, dans sa modélisation de la gestion des processus rédactionnels, Olive (2014) indique qu'un des problèmes restant à aborder est celui de l'unité de préparation du langage. Cette question est d'autant plus délicate dans le cadre d'un texte, où la segmentation en unités n'est pas toujours aisée, Legallois (2006, p. 3) indiquant même que « lorsque la linguistique se donne le texte pour objet, force est d'admettre que la question des unités demeure une énigme insoluble ». Nous essaierons de répondre à cette problématique des unités de production grâce à la description du contenu linguistique des textes produits, mise en relation avec les données temporelles à notre disposition.

Cette thèse s'inscrit donc dans le champ de la « psycholinguistique textuelle », qui, comme le définissent Coirier, Gaonac'h et Passerault (1996, p. 6), consiste en « l'étude des mécanismes par lesquels le sujet humain traite les dispositifs linguistiques, et plus largement les dispositifs langagiers, en vue de produire et/ou d'interpréter une suite cohérente d'énoncés », notre travail portant uniquement sur la production écrite. Dès lors, il apparait nécessaire de se situer à la croisée de différentes disciplines. D'un point de vue des cadres théoriques, il est indispensable de se référer d'une part à la linguistique, qui permet de décrire avec précision en quoi consistent ces « énoncés » et comment en assurer « une suite cohérente », et d'autre part à la psychologie cognitive, dont les modèles pourront expliciter quels

sont les processus engagés lors de la rédaction d'un texte et comment ils sont gérés. La production de tels modèles, ainsi que l'intérêt porté à la manière dont les enfants développent leurs compétences langagières, conduisent naturellement à poser la question du lien entre la psycholinguistique et la didactique (Fayol, 1990; Garcia-Debanc & Fayol, 2013) : tout en ayant conscience de la limite de ces modèles, il peut en effet être intéressant de voir comment ils parviennent à rendre compte de la complexité de l'activité proposée à l'élève, permettant ainsi de mieux comprendre certaines difficultés, et aidant à concevoir des aides pour les activités d'écriture. La complémentarité des disciplines ne se situe pas seulement au niveau des cadres théoriques ou du but de la recherche, mais également de la méthodologie employée. En effet, un paradigme méthodologique employé en psychologie cognitive pour mettre à jour la gestion des processus rédactionnels est l'analyse en temps réel, dont les techniques « visent à pister les processus cognitifs de rédaction pendant leur fonctionnement, ceci afin de décrire leur décours temporel et leurs caractéristiques fonctionnelles » (Olive, 2011, p. 61). Plus particulièrement, une de ces techniques est l'étude des pauses et de la vitesse d'écriture : il est considéré que la variation de ces données chronométriques reflète les différences de cout cognitif, devenant ainsi un indice des processus cognitifs engagés à ce moment de l'écriture (Schilperoord, 2002). Cette utilisation des données chronométriques comme indices de processus sous-jacents nous rapproche de la méthodologie utilisée en génétique textuelle (Plane, Alamargot & Lebrave, 2010), qui peut fournir des indices complémentaires. Dans cette discipline, les chercheurs ont notamment utilisé les brouillons « comme traces de processus, comme inscription matérielle d'événements dont il fallait reconstruire la dynamique en temps réel » (Lebrave & Grésillon, 2009). En outre, l'utilisation d'outils permettant l'enregistrement de l'activité d'écriture permet de compléter les indices disponibles grâce à la « temporalité relative », qui indique la succession temporelle des opérations, et la « temporalité absolue » qui est donnée en référence au temps initial (Leblay, 2009), auxquelles s'ajoutent les durées des pauses et des périodes d'écriture liées à ces révisions (Doquet, 1995).

Le but de notre travail est, tout d'abord, de caractériser la taille des unités de production d'un point de vue développemental. En effet, les études ayant comparé l'unité de production des enfants d'école primaire à celle des adultes (Chanquoy, Foulin & Fayol, 1990; van Hell, Verhoeven & van Bijsterverldt, 2008) ont montré des différences entre les deux types de population ; cependant, elles ne proposent pas de pattern développemental. L'analyse de productions de scripteurs âgés de 10 à 15 ans pourrait donc être pertinente pour mieux comprendre ce développement. De plus, la question des unités étant certainement liée à celle de la maitrise des normes de segmentation, cette période est d'autant plus pertinente à étudier qu'il s'agit du moment où l'usage de la ponctuation devient davantage normé (Fauchart, 2001). Le deuxième objectif de ce travail est d'étudier la gestion temporelle de l'écriture au

niveau des clauses[1] et des phrases. En effet, le travail de Maggio (2013) a permis de caractériser le traitement du mot dans le cadre de l'écriture d'un texte, en liant les données chronométriques (pauses en début de mot, pauses intra-mots et débit d'écriture) aux caractéristiques lexicales et infra-lexicales du mot en cours d'écriture, ainsi que du mot précédent et du mot suivant. Après avoir défini des unités de production possibles, nous souhaitons intégrer les caractéristiques des autres niveaux d'analyse du texte tels que le niveaux syntaxique et le niveau sémantique, pour étudier plus précisément à quel moment les choix sémantiques, syntaxiques et lexicaux concernant ces unités sont effectués. Les données chronométriques seront donc de nouveau utilisées, mais seront cette fois-ci mises en relation avec les caractéristiques linguistiques des unités permettant de rendre compte de leur construction interne et du lien entretenu avec les unités environnantes. Enfin, nous souhaitons compléter notre recherche par les indices fournis par les modifications apportées sur le texte déjà écrit, afin de voir dans quelle mesure il est possible d'utiliser les informations temporelles liées aux révisions pour mieux comprendre la gestion des unités de production.

Notre thèse sera présentée en trois parties. La première partie sera consacrée au cadre théorique. Dans le chapitre 1 seront abordées les contraintes d'ordre linguistique : nous présenterons comment le scripteur doit passer d'une juxtaposition de clauses à un texte cohérent, notamment par le biais de liages sémantique et syntaxique ; les unités de segmentation identifiables lors de l'analyse textuelle seront également présentées, tout en soulignant les difficultés rencontrées par les scripteurs pour segmenter correctement leur texte. Dans le chapitre 2 seront mentionnées les contraintes d'ordre cognitif, avec notamment la présentation de différents modèles, permettant d'identifier les processus rédactionnels et la manière dont ils sont gérés par le scripteur. Enfin, le chapitre 3 conduira vers le propos de notre recherche : les méthodes d'analyses utilisant les données temporelles seront explicitées, puis la problématique de la gestion des unités de production sera introduite, suivie par une revue des différentes études pouvant contribuer à répondre à cette question.

La partie 2 sera constituée d'un seul chapitre, qui présentera la méthodologie générale : y seront décrits les deux corpus utilisés pour notre recherche, les codages généraux portant sur la segmentation des textes et les données temporelles, ainsi que les types d'analyses qui seront utilisés dans notre travail.

La partie 3 présentera les résultats, répartis dans quatre chapitres. Le chapitre 5 s'appuiera sur le corpus développemental, afin d'étudier le développement des unités de production et le lien entre ce développement et la maitrise des normes typographiques de segmentation. Les autres chapitres porteront sur le corpus d'adultes ayant rédigé la suite d'un récit fictif. Le chapitre 6 présentera tout d'abord les données temporelles, en mentionnant quelques différences interindividuelles, puis fera le lien entre les

[1] Nous utiliserons dans notre étude principalement le terme de « clause », que nous définirons plus précisément par la suite, qui correspond globalement au terme de « proposition » souvent employé dans les grammaires traditionnelles françaises ; ces deux expressions seront donc employées indifféremment.

données chronométriques et l'empaquetage syntaxique afin de faire ressortir les unités de production qui semblent pertinentes. La gestion temporelle de ces unités sera analysée dans le chapitre 7, grâce à la mise en relation des données chronométriques avec le contenu linguistique de ces unités. Enfin, le chapitre 8 permettra de compléter ces analyses grâce à l'étude des informations temporelles liées aux révisions présentes dans ce corpus ; en outre, une étude de cas permettra de montrer la complémentarité des indices à notre disposition pour comprendre la gestion des processus rédactionnels et ouvrir de nouvelles pistes de recherche.

Cadre théorique

Chapitre 1 - Produire un texte : la complexité du produit fini

Un des aspects complexes de la rédaction d'un texte tient à l'objet-même qu'il est demandé de produire : un texte, c'est-à-dire une suite cohérente d'énoncés. Afin de produire ce texte, le scripteur doit gérer des contraintes locales liées à sa langue (sur les plans orthographique, lexical, grammatical, sémantico-logique), des contraintes textuelles et des contraintes liées aux pratiques discursives historiquement et socialement réglées (Adam, 1993). En effet, « la compétence textuelle s'appuie sur de nombreuses compétences linguistiques « locales », que l'on peut analyser au niveau phrastique ou interphrastique. Mais elle ne se réduit pas à la somme, même complexe, de ces compétences locales ; et cela tient dans une large mesure au fait que tout texte est caractérisé par un ensemble de structures non linguistiques, structures associées en particulier à des schémas cognitifs sous-jacents » (p. 138).

Le but de notre travail est d'examiner, lorsqu'une personne écrit un texte, quelles sont les unités de production et dans quelle mesure certaines sont produites plus ou moins facilement. Pour cela, il est donc nécessaire de prendre en compte le contenu linguistique qui est produit, lié aux contraintes pesant sur la production textuelle. Nous présenterons donc tout d'abord comment le scripteur est amené, à partir de la succession d'unités minimales, à produire un tout cohérent. Cela nous permettra d'expliciter un certain nombre d'éléments linguistiques qui peuvent être analysés pour étudier le contenu linguistique du texte, dans le but d'expliquer la gestion de la production écrite ; nous présenterons les différences de contenu linguistique identifiables en fonction du type de texte et de l'âge des scripteurs, deux variables qui seront à prendre en compte dans nos études. Enfin, puisque notre travail porte sur les « unités », nous nous interrogerons sur celles-ci : nous verrons quelles sont les unités identifiables par le chercheur qui étudie le texte, mais nous soulèverons les problématiques rencontrées par le scripteur lorsqu'il doit segmenter le texte ; cela nous permettra de poser la question du rapport entre unité observée et unité de production.

1. D'une juxtaposition de clauses à un texte cohérent

L'unité qui servira de base à notre travail est la « clause ». Nous présenterons tout d'abord la définition que nous retenons pour cette unité, en la mettant en regard avec d'autres définitions parfois proposées. Nous expliciterons ensuite la notion de « cohérence », nécessaire pour parler de « texte ». Les outils permettant de garantir la cohérence, à savoir les éléments servant aux liages sémantique et syntaxique, ainsi que la constitution d'un schéma textuel, seront ensuite exposés.

1.1. De la clause au texte

1.1.1. L'unité minimale : la clause

L'unité minimale choisie est la clause, unité à partir de laquelle il est possible d'envisager des assemblages. Nous la définissons à partir des termes de Berman et Slobin (1994, p. 7) : « Tout unité qui contient un prédicat unifié », c'est-à-dire « un prédicat qui exprime une seule situation (activité, évènement, état), incluant des verbes fléchis et non fléchis ainsi que des prédicats adjectivaux »[2]. Le détail de la segmentation en clauses, et donc de ce qui est entendu par « prédicat unifié », sera explicité dans le chapitre « Méthodologie générale ». La clause est donc une unité rectionnelle s'organisant autour d'un prédicat.

Le modèle de Fribourg (Béguelin, 2002, p. 7) définit également la clause au niveau syntaxique comme un « ilot maximal de solidarités rectionnelles », mais insiste sur le plan fonctionnel, puisque la clause accomplit « un acte énonciatif minimal, visant à modifier l'état des représentations partagées par les interlocuteurs ». Cet aspect fonctionnel est également privilégié par Adam (2011a), lorsqu'il indique que la proposition[3], qui est une « micro-unité syntaxique et micro-unité de sens » sert à faire le « lien entre un objet du discours [...] et ce qui est dit à l'aide ou non d'un prédicat verbal ». Ces dernières définitions ne sont pas incompatibles avec celle de Berman et Slobin, bien que nous préférions définir la clause à partir du prédicat. Dans ce sens, notre définition correspond bien à celle de Kintsch (citée dans Coirier et al., 1996, p. 14), pour qui « une proposition est une unité de signification qui contient un ou plusieurs arguments. Ces arguments sont des entités référentielles pouvant correspondre à des êtres, des objets, des idées, d'autres propositions… Des prédicats assignent des propriétés aux arguments, ou définissent la relation entre les arguments ». Ces clauses sont ensuite assemblées en unités

[2] Notre traduction : « any unit that contains a unified predicate » by which we mean « a predicate that expresses a single situation (activity, event, or state), including finite and non finite verbs as well as predicat adjectives »

[3] Pour rappel, le terme de « proposition » utilisé par les chercheurs francophones correspond généralement à ce que nous avons nommé « clause » ; ces deux termes peuvent donc être utilisés de manière équivalente.

plus complexes, par exemple des phrases ; il ne suffit pourtant pas de juxtaposer des phrases pour qu'un « texte » soit produit.

1.1.2. La nécessité de cohérence

Il est indispensable que la production du scripteur soit cohérente, afin qu'elle soit considérée comme un texte, ayant donc une unité de sens, et non comme une simple juxtaposition de phrases. C'est justement parce qu'il y a une cohérence que l'on peut parler de texte, ou de discours[4], et à l'inverse parce qu'une production est présentée comme un texte que l'on cherche la cohérence entre les énoncés : la cohérence est donc constitutive à l'idée de « discours » (Charolles, 2011). Des règles d'ordre linguistique et pragmatique ont pu être énoncées pour expliciter ce qu'est un texte cohérent : les méta-règles de répétition, de progression, de non-contradiction et de relation (Charolles, 1978). Ainsi, le scripteur doit maintenir un thème tout en le développant ; les nouveaux éléments introduits ne peuvent contredire d'un point de vue sémantique ou pragmatique ce qui a été dit ; tous ces éléments doivent être reliés.

À la notion de « cohérence » est souvent associée celle de « cohésion » (il est notamment possible de se référer à Adam, 2011; Coirier et al., 1996; Sanders & Pander Maat, 2006) ; il convient donc de distinguer ces deux termes. Suite à ce qu'il vient d'être précisé concernant la cohérence, on peut considérer que celle-ci implique un phénomène cognitif et se situe du côté du destinataire : le fait qu'un texte soit cohérent signifie que le destinataire est capable de se construire un modèle mental du message exprimé par le locuteur[5] ; il s'agit donc de la représentation cognitive associée au texte. La cohésion, quant à elle, concerne les aspects linguistiques du texte. Dans la relation cohérence – cohésion, il est fréquemment dit que la cohésion est l'apparition « en surface » de la cohérence, c'est-à-dire qu'il s'agit des outils linguistiques utilisés pour manifester la cohérence d'un texte (Favart, 2005).

Cependant, les liens entre ces deux notions ne sont pas si clairs, et il est possible de se demander dans quelle mesure la cohésion est un critère suffisant pour assurer la cohérence, voire même si c'est un critère nécessaire (Salles, 2006). Ainsi, plusieurs études ont cherché à montrer une corrélation entre la cohésion, donc les outils linguistiques utilisés, et la cohérence. La cohésion peut être décrite objectivement par la description de nombreux éléments linguistiques du texte, alors que la cohérence est appréhendée grâce au ressenti du destinataire, notamment à travers certaines échelles de qualité qui prennent en compte une variable de type « structuration du contenu » (Beauvais, Olive & Passerault, 2011) ou « disposition et présentation logique des idées » (Ferrari, Bouffard & Rainville, 1998). Or, si

[4] Dans notre travail, les termes « textes » et « discours » sont utilisés de manière équivalente pour désigner des productions monologiques constituant un tout ; la notion de « discours » est parfois privilégiée pour désigner des productions orales.

[5] Les études de psycholinguistique textuelle portent aussi bien sur les modalités orales qu'écrites. Notre travail ne portant que sur l'écrit, il arrivera que nous utilisions les termes « sciprteur » et « lecteur », plus spécifiques, à la place de « locuteur » et « destinataire ».

la cohérence apparait comme un facteur important de la qualité des textes (McCulley, 1985; McNamara, Crossley & McCarthy, 2010), l'utilisation d'outils cohésifs ne l'est pas forcément (Tierney & Mosenthal, 1983). Un outil informatique particulièrement intéressant pour ces questions a été développé pour les données anglaises : il s'agit de Coh-Metrix (Graesser & McNamara, 2004), qui fournit plus de 200 variables textuelles, dont notamment de nombreux outils cohésifs (e.g. connecteurs, opérateurs logiques, références anaphoriques). Il est ressorti des études ayant utilisé Coh-Metrix que les productions de bonne qualité se distinguaient par un texte plus complexe et sophistiqué (caractérisé par une plus grande complexité syntaxique, plus de diversité lexicale et des mots moins fréquents), mais non par l'utilisation de davantage d'outils de cohésion (Crossley & McNamara, 2011a, 2011b) ; bien plus, la corrélation entre les outils cohésifs et la cohérence est négative, hormis pour les conjonctions de coordination (Crossley & McNamara, 2010). Ce résultat est à mettre en relation avec notamment l'état de connaissances du scripteur, du destinataire, et de ce que le scripteur sait du destinataire : si chacun a de bonnes connaissances externes, et que le locuteur identifie bien ce qu'il est suffisant de dire pour que le message soit bien compris, il n'est pas nécessaire d'avoir recours à un trop grand nombre d'outils cohésifs.

Même si le lien n'est pas toujours bien établi entre la quantité d'outils de cohésion utilisés et la qualité de la cohérence du texte, le respect des règles de cohérence énoncées ci-dessus prend comme support les moyens linguistiques disponibles. Parmi les règles de cohérence, nous nous intéresserons plus particulièrement aux méta-règles de répétition, progression et relation. Plusieurs catégorisations des outils de cohésion ont été proposées. Les études portant sur la cohésion se réfèrent souvent aux travaux de Halliday & Hasan (1976) qui distinguent la référence, la substitution, l'ellipse, les connecteurs et la cohésion lexicale, auxquels ils ajoutent les signes de ponctuation (à l'écrit) ou l'intonation (à l'oral). Adam (2011a) identifie quant à lui plusieurs manières de lier les propositions : le liage du signifié, le liage du signifiant (par le biais des parallélismes, allitérations, rimes), les implications (ellipses, implicites), les connexions et les séquences d'acte de discours. Les implications et les séquences d'acte de discours concernent plutôt l'aspect pragmatique ; le liage du signifiant est particulièrement utilisé dans certains genres discursifs tels que la poésie. Comme, au niveau de la cohérence, nous nous intéressons essentiellement aux méta-règles de répétition, progression et relation, nous examinerons le liage « sémantique » (notamment la continuité référentielle et les collocations) et le liage « syntaxique » (c'est-à-dire les connecteurs et marqueurs discursifs, ainsi que l'empaquetage syntaxique). La cohérence impliquant la constitution d'un « modèle mental », nous aborderons ensuite la notion de « schéma textuel », utile pour construire un tel modèle.

1.2. L'assemblage des clauses

1.2.1. Liage sémantique

Le liage sémantique concerne ce que Halliday et Hasan (1976) nommaient « référence » et « cohésion lexicale ». Les éléments du texte peuvent ainsi être liés entre eux soit parce qu'ils désignent un même référent, soit parce qu'ils appartiennent au même champ lexical.

La première règle de cohérence concerne la répétition : il est en effet attendu que certains référents soient repris dans le texte, constituant des « chaines de référence » (Schnedecker & Landragin, 2014). La co-référence correspond donc à une « relation d'identité référentielle entre deux ou plusieurs signes sémantiquement interprétables indépendamment l'un de l'autre » (Adam, 2011a, p. 104). Cette co-référence est anaphorique si elle reprend un référent présent dans le co-texte « gauche », c'est-à-dire le texte déjà écrit ou lu, et elle est cataphorique si elle renvoie à un élément qui sera indiqué dans le co-texte « droit ». Elle peut se faire grâce à des pronoms (e.g. personnels, possessifs) ou des syntagmes nominaux. Dans le cas d'une reprise sous forme lexicale, une certaine progression peut apparaitre, par la mention d'une propriété du référent par exemple (e.g. « la jolie fille », « la vendeuse »). Il ne s'agit donc pas nécessairement de répétition exacte, le scripteur pouvant avoir recours à la synonymie ou à l'hyperonymie. Généralement, l'utilisation correcte de la continuité référentielle aide au traitement de l'information et à la mémorisation du texte (Garnham, Oakhill & Johnson-Laird, 1982).

La deuxième règle de cohérence concerne la progression. Nous venons d'indiquer que celle-ci pouvait provenir du choix d'un item lexical différent pour désigner un référent déjà introduit. Il est aussi possible de « progresser » dans le texte grâce à l'introduction de nouveaux référents. Plusieurs auteurs mentionnent la distinction entre « ancienne » et « nouvelle » information (Halliday & Hasan, 1976; Haviland & Clark, 1974). Cependant, la définition d'« ancienne » information peut varier : il peut s'agir d'un référent explicitement mentionné dans le texte déjà écrit, ou d'un référent qu'il est possible de retrouver dans le discours. Ainsi, il y aurait un statut intermédiaire d'information « accessible » ou « inférable » (Chafe, 1994) : on peut alors parler de continuum entre « nouvelle » et « ancienne » information, Prince (1981) parlant de « hiérarchie de familiarité » et Givon (1983) d'« échelle d'accessibilité des référents ». Le fait qu'un référent soit ancien, plus ou moins accessible ou complètement nouveau va normalement influencer les choix linguistiques du scripteur. En effet, une ancienne information sera généralement reprise sous forme de pronom et sera de préférence placée comme sujet de la phrase ; des déterminants comme les définis ou les démonstratifs seront plutôt réservés aux référents accessibles ; un syntagme nominal indéfini sera en principe nécessaire pour introduire un référent totalement nouveau (Allen, Skarabela & Hughes, 2008; Ariel, 1990, 2001; Mazur-Palandre, 2015). Ainsi, la progression du texte n'est pas si facile à respecter car des choix linguistiques pertinents doivent être effectués, et ces choix étant notamment liés à ce que le locuteur considère comme accessible par le destinataire.

Nous avons évoqué la notion d'« accessibilité » pour des référents qui n'étaient pas explicitement mentionnés auparavant. Un nouvel élément peut être facilement accessible s'il est induit par la situation ou s'il est sémantiquement relié à un autre terme : par exemple, le fait de parler de « maison » rend accessible l'élément « toit » ou « chambre ». Dans cet exemple, les termes du champ lexical « maison » seront donc assez facilement accessibles. En outre, le fait d'avoir des termes du même champ lexical, ou la présence de termes reliés sémantiquement (e.g. antonymes, synonymes) renforcent la cohésion lexicale (Adam, 2011a).

1.2.2. Liage syntaxique

Une des règles de cohérence concerne l'importance des liaisons entre les propositions : c'est ce que Charolles (2011) appelait « cohérence relationnelle ». Il est en effet important de lier les clauses entre elles, puis les unités supra-clausales (par exemple les phrases) entre elles, afin de passer d'une juxtaposition de phrases sans liens à un texte. En outre, le fait d'utiliser des connecteurs permet de linéariser le texte : il n'est en effet pas toujours aisé de passer d'une représentation conceptuelle parfois hiérarchique, multidimensionnelle, à un texte linéaire (Fayol, 1989). Cela est notamment rendu possible par l'explicitation des liens entre les propositions.

1.2.2.1. Les connecteurs

Ces liens entre les propositions peuvent être exprimés par des connecteurs qui indiquent le degré et la nature de cette relation, et qui permettent de structurer le texte. Ainsi, Halliday et Hasan (1976) distinguent les connecteurs spécifiant une relation causale (« parce que », « car »), temporelle (« puis », « ensuite »), additive (« par ailleurs », « et ») ou adversative (« mais », « cependant »). Il est possible d'y ajouter des connecteurs exprimant une relation de conséquence (« donc », « par conséquent »), de but (« afin de », « pour »), d'exemplification (« par exemple », « ainsi »), ainsi que les marqueurs d'intégration linéaire qui peuvent soit ouvrir une série (« tout d'abord », « d'une part »), soit la clore (« d'autre part », « enfin »). Un point d'attention doit être porté au niveau de la sémantique de ces connecteurs. En effet, certains connecteurs sont polysémiques (Favart, 2005) : par exemple, le connecteur « et » peut exprimer, suivant les situations, une relation temporelle, causale ou additive (Favart & Passerault, 1999).

Ces connecteurs sont généralement des conjonctions ou des adverbes. Parmi les conjonctions, nous pouvons distinguer les conjonctions de coordination (e.g. « et », « mais ») et celles de subordination (e.g. « puisque », « alors que »). La relation entre les deux propositions est donc manifestée d'une part par le sens du connecteur, et d'autre part par la construction syntaxique qui en découle. En effet, le fait de choisir une conjonction de subordination plutôt qu'une conjonction de coordination ou qu'un adverbe renforce la dépendance entre les deux propositions, comme nous allons le montrer à travers l'étude de l'empaquetage syntaxique.

Fayol (1986) propose qu'en réalité deux sous-systèmes servent à indiquer les relations interpropositionnelles : celui des connecteurs et celui de la ponctuation. Ces deux sous-systèmes peuvent d'ailleurs interagir, puisque la redondance d'une conjonction et d'une ponctuation forte peut permettre une mise en relief de l'énoncé, et que par ailleurs la présence d'une marque tend généralement à supprimer l'autre (Fayol, 1996). La question de la ponctuation sera abordée plus en détail ultérieurement, lorsque les difficultés concernant la segmentation du texte seront présentées.

1.2.2.2. L'empaquetage syntaxique

Le choix des connecteurs, particulièrement entre plusieurs synonymes dont la nature grammaticale diffère, dépend notamment du lien syntaxique entre les clauses que l'on souhaite exprimer. Ce lien syntaxique peut parfois être motivé par une économie syntagmatique, si le scripteur souhaite réduire la complexité ou la longueur de la phrase, mais il peut parfois rendre visible un lien de dépendance : ainsi, deux concepts dépendants seraient codés linguistiquement de manière plus dépendante (Berman & Slobin, 1994).

On peut distinguer deux principaux types de relation entre les clauses peuvent être distingués : la « parataxe » et l'« hypotaxe » (Koch, 1995). Dans le cas de la parataxe, les deux clauses conservent leur autonomie syntaxique : il peut s'agir soit de juxtaposées, soit de coordonnées. Le terme d'« hypotaxe », quant à lui, est utilisé lorsqu'il y a un rapport de dépendance entre les deux clauses. Il s'agit des clauses subordonnées, qu'elles soient fléchies (relatives, circonstancielles et complétives finies) ou non fléchies (circonstancielles et complétives non finies). Dans notre étude, d'un point de vue terminologique, nous distinguerons ainsi les clauses « indépendantes » (juxtaposées et coordonnées, relation de parataxe) des clauses « subordonnées » (relation d'hypotaxe). Les catégorisations en clauses « juxtaposées » et « coordonnées » peuvent varier. Il est possible de considérer comme « coordonnée » toute clause comportant une conjonction de coordination ou un adverbe de liaison (Akinci, 2005; Gayraud, Jisa & Viguié, 1999; Riegel, Pellat & Rioul, 2002). Une autre option est de distinguer les juxtaposées « marquées » (introduites par un adverbe de liaison) des « non marquées » (Mazur-Palandre, 2007), gardant uniquement le terme de « coordonnée » si une conjonction de coordination est présente[6].

Nous voyons apparaitre ici la notion de « dépendance », caractéristique des subordonnées, ce que Koch (1995, p. 21) explicite lorsqu'il déclare que « dès qu'il y a subordination, il s'établit une relation de détermination ou bien, comme je préfère le dire, de dépendance, en ce sens que la séquence B dépend, d'une manière ou d'une autre, du verbe de la proposition A. » Cependant, si la dépendance peut être

[6] Les adverbes de liaison peuvent se cumuler entre eux, se combiner avec une conjonction antéposée et sont relativement mobiles, ce qui pousse Riegel et a1.(2002) à classer « donc » parmi ces adverbes, et non comme une conjonction de coordination.

définie comme « l'impossibilité pour une clause d'apparaitre en isolation », peu de phrases peuvent être complètement indépendantes au sein d'un texte, ce dernier présupposant cohérence et cohésion. La dépendance serait donc plutôt une caractéristique graduelle. Par exemple, si l'on compare deux constructions subordonnées, la présence d'un verbe non fléchi implique une plus grande dépendance que si le verbe est fléchi. Même les clauses que nous avons qualifiées d'« indépendantes » le sont à des degrés divers : nous avons déjà mentionné la présence possible d'une conjonction ou d'un adverbe de liaison ; de plus, parmi les coordonnées, certaines ont la particularité de comporter une ellipse du sujet, ce qui rend plus dépendante la clause puisqu'il faut se référer au sujet de la clause précédente : une distinction peut donc être faite entre coordonnées avec et sans ellipse (Mazur-Palandre 2007). Ainsi, le degré de dépendance entre les clauses peut non seulement être signalé par le connecteur et la ponctuation utilisés, mais également par les caractéristiques de l'empaquetage syntaxique.

1.3. La constitution d'un schéma textuel

Nous avons donc vu comment, afin de rendre son texte cohérent, le scripteur devait lier les clauses entre elles, que ce soit d'un point de vue sémantique ou syntaxique. Un texte est considéré comme cohérent si le destinataire parvient à reconstruire le modèle mental transmis de manière linéaire par le scripteur. Cette représentation mentale correspond au niveau superstructural§ (Coirier et al., 1996). Il s'agit de ce qui peut être nommé « modèle de situation » (van Dijk & Kintsch, 1983), « schéma textuel » ou « scénario ». Le fait de posséder le modèle mental adapté au texte demandé permet au scripteur d'organiser plus facilement ses idées. Il peut ainsi s'assurer de ne pas oublier ce qui est essentiel dans la structure, puis ajouter les détails suivant le niveau de précision qu'il souhaite offrir. De même, il sera plus facile pour le destinataire de mémoriser les informations si elles permettent de « remplir » le modèle attendu. Kintsch & van Dijk (1978) proposent en effet que pour reconstruire la macrostructure du texte, le destinataire doit arriver à supprimer les détails, ainsi que ce qui semble important au niveau local mais qui ne l'est pas au niveau global. Le fait d'avoir un schéma textuel disponible en mémoire peut alors permettre de sélectionner les informations pertinentes.

Comme nous le détaillerons par la suite, le schéma textuel dépend du type de texte. De nombreux travaux ont cherché à expliciter la macrostructure des textes narratifs. Tout d'abord, Propp (cité dans Coirier et al., 1996) a identifié que, sous la diversité des contes, il existe une structure formelle permanente ; il a ainsi proposé une « morphologie du conte », puis a essayé d'établir des « grammaires d'histoires ». Malgré certaines critiques adressées à ces modèles, le fait qu'une structure stable permette la présence d'un schéma cognitif potentiel de traitement a fait croitre l'intérêt pour la recherche de modèles superstructuraux. Ainsi, des éléments récurrents comme la présence de complications ou de résolution de problèmes sont attendus dans un récit[7] ; nous évoquerons par la suite quelques modèles

[7] Dans notre travail, nous ne ferons pas de distinction entre les termes « récit » et « texte narratif ».

proposés pour rendre compte de ces schémas narratifs (Adam, 2011b; Labov & Waletzky, 1967). Les textes expositifs ont une structure moins contraignante que les textes narratifs (Boscolo, 1990), mais il sera tout de même possible de présenter quelques éléments attendus lors de la production de ce type de texte.

2. Le contenu linguistique du texte

La partie précédente a mis en avant la complexité de ce qu'est un texte. À partir de là, nous pouvons identifier des éléments qu'il est possible d'étudier plus précisément lorsque l'on s'intéresse à la psycholinguistique textuelle, c'est-à-dire lorsque l'on souhaite comprendre les traitements liés à la compréhension ou, dans notre cas, à la production de texte. Ces éléments linguistiques se situent au niveau de la clause, au niveau interclausal, que ce soit d'un point de vue sémantique ou syntaxique, et enfin au niveau du schéma textuel. Les éléments pouvant être analysés ne sont pas présentés de manière exhaustive. Le but est ici de montrer la complexité du texte concernant ces différents niveaux, en lien avec la cohérence recherchée par le scripteur ; un certain nombre de ces éléments linguistiques seront pris en compte dans notre étude. En outre, un de nos corpus est constitué de textes expositifs et narratifs produits par des élèves de 10 à 15 ans ; l'autre corpus contient des textes narratifs rédigés par des adultes. Nous présenterons donc quelques éléments liés aux différences linguistiques en fonction de ces types de texte et de l'âge des scripteurs, qui pourront être utiles pour comprendre certains résultats obtenus par la suite.

2.1. Éléments linguistiques susceptibles d'être analysés

2.1.1. Au niveau de la clause

La notion de « prédicat » est récurrente dans les définitions de la « clause » proposées au début de ce chapitre (pour une présentation des différentes acceptions de cette notion, voir Maillard, 2008). Celui-ci est en effet le noyau de la clause. Il faut cependant être attentif à l'acception que nous en faisons, puisqu'un des problèmes concernant l'utilisation de ce terme est que plusieurs définitions en sont données (Gaatone, 2008). En effet, un premier sens du prédicat est « ce que l'on dit du sujet », le sujet étant « ce dont il s'agit ». À la suite d'Aristote, la grammaire traditionnelle considère ainsi que la phrase est bipartite : ce dont on parle et ce que l'on en dit, soit en termes syntaxiques le « sujet » et le « prédicat », et en termes communicatifs le « thème » et le « rhème » (Combettes, 1986; Riegel et al., 2002) ; le prédicat est alors généralement le « groupe verbal », c'est-à-dire le verbe et ses compléments. Une des limites de cette conception est que l'on peut dire quelque chose, sans forcément parler « de » quelque chose : par exemple, la phrase « il neige » comporte un rhème, mais pas de thème identifiable. Pour notre part, nous utilisons le terme « prédicat » dans le sens donné par la sémantique logique (e.g. Gross, 1981), le prédicat étant le noyau de la clause, par rapport auquel toutes les fonctions se définissent. Les « arguments » régis par le prédicat sont donc son sujet et ses différents compléments (e.g. complément d'objet direct). Ainsi, une fois son prédicat défini, le scripteur est capable de construire le « squelette » de sa clause à partir de la valence du prédicat, c'est-à-dire du nombre et du type d'éléments régis par ce noyau.

D'un point de vue sémantique, il est justement possible de donner des informations au sujet du prédicat. Les grammaires traditionnelles distinguent couramment les « verbes d'état », qui peuvent être remplacés par « être » (e.g. sembler, paraitre) et les « verbes d'action ». Cette catégorisation peut cependant être affinée. C'est ce qu'ont notamment entrepris Ghiglione, Bromberg, Friemel, Kekenbosch et Verstiggel (1990). Ils distinguent ainsi les verbes « factifs », exprimant une action, les verbes « déclaratifs » et les verbes « statifs », ces derniers indiquant soit un état, soit une possession. Plusieurs descripteurs ont ensuite été utilisés pour obtenir des sous-catégories : apparaissent ainsi les factifs « finalisés » et « non finalisés », les déclaratifs « d'expression », « d'effectuation » ou de « cognition », et les statifs « expérientiels » et « non expérientiels ». Plusieurs critères entrent donc dans cette catégorisation, comme les participants ou la constitution temporelle. L'ajout d'autres critères pourrait alors modifier cette classification. Ainsi, en prenant en compte le rapport à une entité externe, Kekenbosch, Meunier, Richard, Desclés et Flageul (1998) identifient des classes telles que « possession ou transfert de possession », « construction d'une nouvelle entité par assemblage ou modification d'une propriété d'une entité », « atteinte à l'intégrité physique d'une entité ». Aurnargue (2008), quant à lui, s'intéresse particulièrement au déplacement, et distingue les verbes de déplacement, qui indiquent un changement d'emplacement par rapport au cadre de référence terrestre (e.g. avancer, grimper, marcher), et les verbes de changement de posture, qui se réfèrent à des mouvements circonscrits au cadre que détermine l'entité faisant l'action (e.g. s'assoir, se recroqueviller). Enfin, une catégorie spécifique a été proposée par Martin (2007) : il s'agit des verbes « pseudo-statifs », qui sont des « verbes agentifs ou dynamiques utilisés pour décrire un état » (p. 69) : c'est par exemple le cas du verbe « barrer » dans « Une voiture barrait la route » ; le sujet est généralement inanimé. Ces quelques propositions de classification donnent des idées sur la manière dont peut être décrit le prédicat d'un point de vue sémantique. Nous avons vu que la classification dépendait des critères utilisés. Or, Escarabajal et Kekenbosch (1998) ont montré que les verbes et dérivés verbaux étaient davantage regroupés d'après des critères extrinsèques (liés à la situation) que d'après des propriétés intrinsèques. Nous pourrons donc proposer une classification sémantique des verbes en fonction du contenu des textes demandés.

Par ailleurs, d'autres éléments sémantiques pourraient être relevés concernant les autres constituants de la clause. Ainsi, dans un récit qui se caractérise par la présence d'un « héros » ou « personnage central », on peut chercher à savoir si celui-ci est mentionné dans la clause, et si d'autres personnages sont introduits. Suivant le type de texte demandé, on peut également retrouver des informations sur le cadre du récit, c'est-à-dire le lieu et le temps, ou sur les motivations et sentiments éprouvés par les personnages, éléments essentiels pour susciter l'intérêt du lecteur (Coirier et al., 1996). On peut alors s'aider des catégorisations sémantiques proposées concernant les connecteurs pour affiner la description de certains constituants : par exemple, les connecteurs temporels peuvent indiquer la succession, la rupture ou la simultanéité (Paolacci & Favart, 2010).

Au niveau syntaxique, nous avons indiqué que le prédicat était caractérisé par sa valence : cela permet de déterminer quels sont les constituants obligatoires d'un point de vue syntaxique. En outre, il est possible d'indiquer si, au sein de la clause, certains éléments sont obligatoires pour restreindre le référent d'un point de vue communicatif (Riegel et a1., 2002). À l'inverse, certains éléments sont facultatifs, ce qui peut être pris en compte par l'étude des expansions, notamment les compléments de nom nominaux, les adjectifs ou certains adverbes (Myhill, 2009). La longueur des clauses peut être également signalée, ce qui correspond à la « longueur moyenne des énoncés » souvent étudiée en psycholinguistique (e.g. Johansson, 2009). Par ailleurs, les formes verbales peuvent être relevées, par exemple pour voir si le verbe de la subordonnée est fléchi ou non-fléchi (Gayraud, 2001), ou si la voix passive est utilisée (Jisa, Reilly, Verhoeven, Baruch & Rosado, 2002).

Concernant les aspects lexicaux, on peut identifier la longueur du mot, la fréquence des lexèmes et des lemmes, ainsi que la taille et la fréquence du voisinage phonographique (Maggio, 2013). Les catégories grammaticales peuvent être relevées, afin de connaitre par exemple la densité lexicale de la clause (Bertrand & Dumoulin, 2013). La description peut même aller jusqu'à un niveau infra-lexical, en indiquant la consistance orthographique des mots, c'est-à-dire le niveau de variabilité des transcriptions d'un même son, ou la fréquence syllabique (Maggio, 2013).

2.1.2. Concernant les liens entre les clauses

Lorsque le scripteur produit un texte, le contenu linguistique doit être lié à ce qui l'entoure. Au niveau sémantique, il est possible de repérer les chaines référentielles, et d'indiquer au niveau lexical le type de reprise : par exemple, s'il y a une répétition du même item lexical, ou l'insertion d'un nouvel item. Cela peut permettre de repérer sous quelle forme lexicale et en quelle position sont insérées les anciennes informations (Mazur-Palandre, 2009) ; on peut plus particulièrement regarder si le sujet est maintenu d'une clause à l'autre ou d'une phrase à l'autre, et si oui sous quelle forme. Le lien sémantique entre les clauses peut également être indiqué par une valeur issue de l'analyse sémantique latente (ASL) : la similarité sémantique entre des unités est donnée grâce à la construction d'un espace sémantique de très grande dimension basé sur l'analyse des cooccurrences dans un corpus de texte (Foltz & Landauer, 2003). Une corrélation négative entre l'indice de l'ASL et la qualité des textes a été observée, sans doute parce qu'une importante similarité sémantique indique une faible diversité lexicale (Bestgen, 2012). Enfin, la récurrence de certains thèmes peut être repérée grâce à l'identification des principaux champs lexicaux des textes. Ces champs lexicaux peuvent permettre en outre d'identifier les référents qui apparaissent pour la première fois dans le texte mais qui étaient facilement accessibles, affinant ainsi l'échelle d'accessibilité des référents (Ariel, 2001).

La continuité syntaxique peut, elle, être étudiée à travers les connecteurs utilisés, notamment leur fréquence d'utilisation et leur classification sémantique (Favart & Passerault, 1999). Les liens syntaxiques entre les clauses (i.e. juxtaposition, coordination, subordination) peuvent également être

précisés, en relevant aussi les constructions particulières comme l'enchâssement des clauses (Nir, 2008). Le relevé des clauses subordonnées peut permettre de calculer la « complexité syntaxique » entendue comme la proportion de clauses subordonnées (Johansson, 2009). Enfin, le degré de dépendance entre les clauses peut être affiné en fonction de la nécessité de la présence de certaines clauses : nécessité syntaxique (e.g. clauses complétives), nécessité pour définir le référent (e.g. relatives restrictives), nécessité communicative (e.g. certaines clauses circonstancielles).

2.1.3. Concernant le schéma textuel

Certaines caractéristiques concernent le texte dans son ensemble. Tout d'abord, il est possible d'identifier les différents schémas textuels : nous donnerons des détails sur la structure des textes narratifs et expositifs, qui apparaitront dans nos corpus. Nous indiquerons ensuite deux éléments liés aux prédicats qui servent à construire la trame du texte et à mettre certains éléments en relief, notamment dans le cas de textes narratifs : la temporalité du prédicat et la distinction entre premier et deuxième plan. Enfin, il est possible de relever des éléments concernant le rapport du locuteur à son discours, quels que soient la modalité et le type de texte, ce qui est explicité par les caractéristiques de la « position discursive » (discours stance en anglais, Berman, 2005).

2.1.3.1. L'identification de différents schémas textuels

Les textes ont été classés dans des typologies sur la base de différents critères. Werlich (cité dans Coirier et al., 1996) a distingué cinq types de textes en se fondant sur les structures cognitives mises en jeu : les types descriptif, narratif, expositif, argumentatif et instructif. Les textes de nos corpus seront soit de type narratif, soit expositif. Les textes narratifs sont plutôt liés à la perception du temps, et se caractérisent par la présence d'un personnage principal et d'une progression temporelle ; les textes expositifs, organisés davantage autour de thèmes que de personnages, présentent des liens logiques et thématiques dans le but de synthétiser des représentations conceptuelles. La description de ces schémas textuels permet de présenter des textes prototypiques, c'est-à-dire ce qui est attendu au minimum du scripteur pour répondre à la consigne. Cependant, Adam (2011b) souligne qu'au sein d'un texte peuvent être entremêlés différents types de séquences : par exemple, le locuteur peut interrompre une séquence narrative par une séquence descriptive, avant de reprendre sa narration. Dans les textes assez courts (maximum une page) demandés à nos scripteurs, une seule séquence est attendue par texte.

Le texte narratif

Le texte narratif se caractérise par une histoire relatant de manière logique et chronologique des évènements qui sont soit causés soit vécus par des personnages. Adam (1985) donne une définition du récit qui reprend les règles de cohérence précédemment évoquées : un texte peut être qualifié de récit s'il y a « la permanence d'au moins un rôle » afin de garantir la cohérence logico-sémantique, si les

propositions narratives sont liées par « des rapports de condition-conséquence » et si l'on note la présence d'une conclusion « permettant de saisir ensemble les évènements successifs et de dégager le sens global du récit » (Adam, 1985, p. 40). On peut distinguer plusieurs types spécifiques de textes narratifs, notamment dans le rapport entretenu avec les personnages et l'histoire. Ainsi, un type spécifique est la narration personnelle : il s'agit de décrire une expérience passée ; les propositions linguistiques relatent donc des évènements qui se sont réellement déroulés (Labov & Waletzky, 1967). En revanche, d'autres narrations relatent l'histoire d'une tierce personne, voire de personnages fictifs : c'est par exemple le cas des contes ou des mythes.

Un texte narratif raconte les évènements liés à un personnage, à travers une série de propositions chronologiquement reliées. Même si la structure temporelle peut parfois varier, en racontant par exemple un fait plus ancien puis en revenant au moment de l'évènement narré, il est possible de reconstituer facilement l'enchainement chronologique des évènements, assez proche de l'enchainement séquentiel des propositions. Concernant la macrostructure du texte, plusieurs chercheurs ont essayé de formaliser la structure narrative. Ainsi, en se basant sur des versions orales d'expériences personnelles, Labov et Waletsky (1967) distinguent cinq parties : l'orientation, la complication, l'évaluation, le résultat, la coda. Parmi ces parties, seuls la complication et le résultat « narrent » des évènements ; l'évaluation permet une transition entre les deux parties narratives, en faisant ressortir le moment crucial de la narration, qui montre justement l'intérêt de cette narration. Les parties extrêmes sont avant tout définies pour les cas des récits d'expériences, pour lesquels le locuteur narre un fait passé : il est alors nécessaire de faire un lien entre les faits racontés et la situation d'énonciation ; l'orientation permet donc de faire le lien entre le contexte d'énonciation et le monde où la narration se situe, et la coda ramène de ce « monde » narratif au contexte d'énonciation. Dans des récits où il n'est pas du tout fait référence à la situation d'énonciation, l'orientation permet de présenter les actants du récit (personnages, lieux, etc.), et la coda récapitule le récit, et peut par exemple être l'occasion de proposer une morale. Un autre schéma textuel, qui convient aussi bien aux récits d'expériences qu'aux récits fictifs coupés de la situation d'énonciation, est proposé par Adam (2011b), qui essaie de prendre en compte les différentes propositions faites par les auteurs avant lui, notamment dans le cadre de la psychologie cognitive et de la psycholinguistique textuelle (e.g. Fayol, 1985; Kintsch & Van Dijk, 1975). L'adaptation de son modèle donne le schéma présenté par la Figure 1.

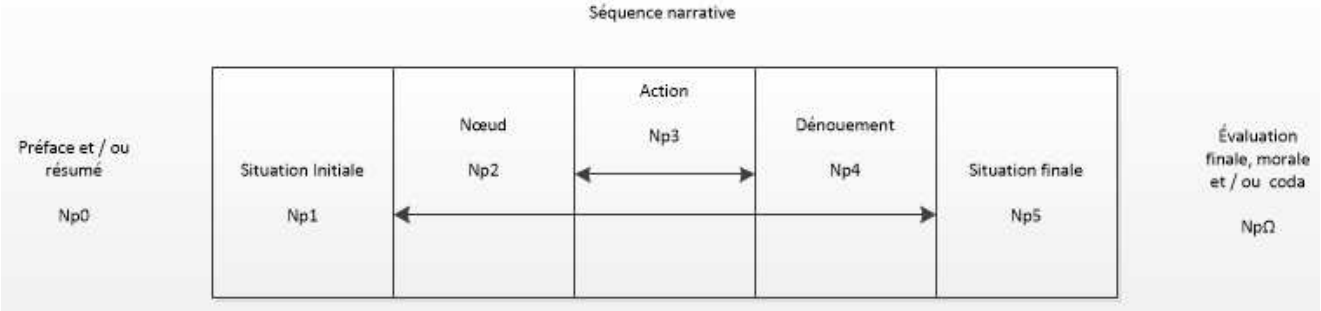


Figure 1. Schématisation des composantes d'une séquence narrative d'après Adam (2011)[8]

La situation initiale (Np1) indique les éléments nécessaires pour débuter le récit, notamment les informations spatio-temporelles et les relations entre les personnages ; l'état décrit dans la situation initiale est interrompu par le nœud déclencheur (Np2) ; les actions sont ensuite racontées (Np3), puis le dénouement (Np4) conduit vers la situation finale (Np5). Adam raisonne dans ce schéma en termes d'organisation autour d'un « procès » central : le procès se déroule pendant Np3, il débute en Np2 et se termine en Np4 ; Np1 décrit l'état avant le procès et Np5 l'état après le procès. Chacune de ces parties (Np) est nommée dans ce modèle « macroproposition ». Deux autres macropropositions d'ordre plutôt pragmatique peuvent être ajoutées : la préface, qui introduit la narration (comme l'« orientation » du modèle de Labov) et la coda qui « ramène » du monde narré, et fournit parfois une morale en faisant une évaluation de l'ensemble de l'histoire. L'ensemble de ces macropropositions forme la séquence narrative.

En outre, il serait possible de regarder plus en détail la fin des récits, par exemple grâce à la typologie des clôtures narratives (Marti, 1999). Ce dernier propose de distinguer les clôtures au niveau logique (à quelle macroproposition appartient-elle), au niveau narratif (fermante ou ouvrante), sémantique (isotope ou non isotope) et socio-historique (stéréotypée ou non stéréotypée). Enfin, la macrostructure du texte narratif peut également s'appuyer sur la notion de « script » (Abelson, 1981; Schank & Abelson, 1977), qui désigne la représentation conceptuelle que possèdent des personnes concernant des séquences évènementielles stéréotypées. L'exemple de ces auteurs concernait les évènements liés à une scène de restaurant : se signaler auprès du serveur, aller à la table, demander le menu, etc. La présence de tels scripts peut faciliter la compréhension, puisque le destinataire connait a priori le déroulement chronologique des évènements ; il peut aussi aider en production, puisque le locuteur sait quelles étapes il doit mentionner, et peut ensuite choisir de détailler certains évènements. Dans notre étude, le récit fictif demandé à nos participants était un récit d'évasion. Bien que le script pour ce type de séquence évènementielle soit moins stéréotypé que le fait d'aller au restaurant, certaines étapes peuvent cependant être communes pour répondre à ce scénario : maitriser le gardien, s'enfuir à travers la prison en déjouant

[8] Notre traduction

la surveillance des gardiens, trouver une ouverture pour s'enfuir de la prison, partir et retrouver d'éventuels complices.

Le texte expositif

Le but du texte expositif est d'être informatif et explicatif à propos d'un sujet, qu'il s'agisse d'informations factuelles ou d'idées théoriques (Boscolo, 1990). La structure n'est donc pas chronologique puisque le sujet ne concerne pas les évènements liés à un personnage, mais un thème qui doit être présenté avec une articulation logique.

La structure des textes expositifs est caractérisée de façon beaucoup moins claire que celle, par exemple, des textes narratifs (Boscolo, 1990). En effet, ce type de texte doit contenir une articulation logique des idées, mais il n'est pas attendu qu'il suive une structure chronologique. Cependant, l'étude de textes expositifs peut faire ressortir trois constituants : l'introduction, permettant de présenter de façon générale le thème de la production ; le corps du texte, où les différentes idées sont développées ; la conclusion, qui permet de terminer le texte en revenant à la thématique générale (Katzenberger, 2004, 2005). Au sein du corps du texte, le scripteur est amené à présenter des sous-thèmes, à les développer, puis à les résumer, des transitions devant être fournies pour lier ces sous-thèmes. La connaissance de cette macrostructure attendue peut faciliter la tâche d'organisation du scripteur ; elle reste cependant toujours difficile, puisque l'articulation logique des thématiques n'est pas aisée, et que c'est au scripteur de déterminer quelles thématiques seront sélectionnées pour développer le thème général, et dans quelle mesure elles seront détaillées.

2.1.3.2. Construction de la trame narrative à partir de caractéristiques du prédicat

La trame temporelle

Un élément qui peut apparaitre comme facilement descriptible concernant le verbe est son temps. Comme nous ne limitons pas le prédicat à un verbe, nous parlerons plutôt de « temporalité du prédicat » ; nous nous appuierons cependant sur ce qui a été étudié au sujet du verbe pour caractériser cette temporalité.

L'étude du temps verbal dans un cadre dépassant la phrase a été renouvelée par la distinction opérée par Benveniste (1966) entre l'« énonciation de discours » et l'« énonciation historique ». Il propose en effet de se référer à la situation énonciative : si dans l'énoncé des références sont faites au moment de l'énonciation, il s'agit d'une « énonciation de discours » ; si l'énoncé est « coupé » de la situation d'énonciation, il s'agit d'« énonciation historique », aussi appelée « récit ». Au premier type d'énonciation correspondent principalement les temps comme le présent, le passé composé, le futur simple ou périphrastique ; les temps du récit sont plutôt le passé simple, l'imparfait, le plus que parfait et le passé antérieur (pour exprimer l'antériorité), et le conditionnel présent (pour exprimer la

postériorité). Les termes « discours » et « récit » pouvant être ambigus, Maingueneau (1993) préfère parler d'« embrayage » lorsqu'il y a une référence au moment de l'énonciation. Ainsi, le « plan embrayé » correspond à « l'énonciation du discours » ; il peut par exemple s'agir d'une conversation, ou d'une personne racontant ce qu'elle a fait la veille en se référant au moment présent ; le « plan non embrayé » correspond au « récit », il peut par exemple s'agir d'un conte ou d'une vérité générale.

La présentation des temps typiques de l'une ou l'autre énonciation montre qu'il faut également raisonner en termes d'antériorité, simultanéité et postériorité par rapport à un évènement. C'est pourquoi il est plus précis de se référer à la proposition de Reichenbach (cité dans Molendijk, 1985), qui distingue le moment de l'énonciation (S), le moment occupé par le fait rapporté (E) et le point référentiel (R) (pour une proposition similaire, voir Fiorin, 1998). Par exemple, si l'on veut parler d'un fait (E) situé dans le futur par rapport à un point référentiel (R), mais que R est passé par rapport au moment de l'énonciation (S), on utilisera un conditionnel présent (ex : « Il se décida. Le lendemain, il **irait** la retrouver »). Le problème est que ce « point référentiel » évolue et qu'il n'est pas bien défini, donc parfois utilisé de manière intuitive. Certains auteurs se sont cependant appliqués à donner de nombreux critères permettant de le définir de manière plus précise que ce qui avait été proposé par Reichenbach. Cette approche de la temporalité a notamment été utilisée par Molendijk (1985, 1990) pour expliquer la distinction entre imparfait et passé simple.

Distinction entre 1er plan et arrière-plan

Lors de l'écriture d'un texte, particulièrement narratif, le scripteur est amené à « donner du relief à [son] texte en projetant au premier plan certains contenus et en en repoussant d'autres dans l'ombre de l'arrière-plan » (Weinrich, 1973, p. 107). Une difficulté est qu'il n'y a pas de règles absolue pour distinguer les deux plans, puisqu'aucun critère n'est ni nécessaire ni suffisant, et qu'il s'agit donc plutôt d'un continuum (Molendijk, 1990; Weinrich, 1973). Nous pouvons cependant présenter certains critères qui sont généralement utilisés pour distinguer les deux plans.

Tout d'abord, il parait important que les séquences constituant le 1er plan soient temporellement ordonnées. Il y a une nécessité de causalité, puisque l'action présentée sert à faire avancer l'histoire ; l'assertion peut être parfois négative, si elle indique ce qui aurait dû arriver, l'absence de cette action étant à l'origine de la suite (Labov & Waletzky, 1967). Ainsi, les évènements relatés au 1er plan sont indispensables à la narration, donc réels, l'ordre temporel de l'enchainement de ces propositions devant correspondre au déroulement des faits dans la réalité (Rabatel, 2004). Dès lors, il apparait que le 1er plan concerne l'action principale du récit, c'est-à-dire ce que retiendrait un compte-rendu factuel (Weinrich, 1973), le prédicat étant donc généralement un verbe d'action présentant un évènement dynamique, et non une situation statique.

Le fait de mettre un évènement au 1er plan conduit à privilégier certains choix linguistiques. Ainsi, il est couramment dit que l'imparfait est le temps prototypique de l'arrière-plan, alors que le passé simple est celui du 1er plan (Molendijk, 1990; Rabatel, 2004; Weinrich, 1973). Cette proposition n'est pas toujours vérifiée, puisque même si Labov (1967) remarque qu'il y a davantage de passé simple dans les parties concernant le développement de l'action que dans celles concernant l'orientation ou l'évaluation (début et fin du récit), il relève tout de même la présence de nombreux verbes à l'imparfait au point central de la narration. Or, si ces évènements sont enlevés, il y a un « trou » dans la narration, ce qui les rend donc indispensables, et le fait de les remplacer par un passé simple amoindrirait l'intensité de ce qui est narré ; il est donc délicat de les considérer comme de l'arrière-plan. D'après Weinrich (1973), d'autres temps seraient également privilégiés pour les évènements d'arrière-plan : il s'agit des formes semi-finies, comme l'infinitif, le participe présent ou passé, ou le gérondif, ces formes étant généralement accompagnées d'un verbe fini. En outre, ce même auteur suggère que les propositions subordonnées servent généralement à indiquer des évènements d'arrière-plan.

Ainsi, de manière prototypique, les prédicats de 1er plan sont des verbes d'action, perfectifs, situés dans la proposition principale, alors que les prédicats d'arrière-plan sont des verbes d'état, duratifs, imperfectifs et situés dans la proposition subordonnée (Fleischman, 1985). Cependant, il est difficile de prévoir ce que le scripteur décidera de mettre au 1er plan et à l'arrière-plan ; en outre, il est délicat de ne se baser que sur les choix syntaxiques pour distinguer entre ces deux plans. Le choix pour le scripteur est pourtant important : à la lecture, le récit en « tire son tempo spécifique » (Weinrich, 1973, p. 117) ; il sera possible de se demander dans quelle mesure le tempo de l'écriture est également affecté par ce choix.

2.1.3.3. La position discursive

Une notion touchant au texte dans son ensemble et se référant à diverses composantes (syntaxiques, sémantiques et pragmatiques) est l'analyse du rapport du locuteur à son texte : quel est le degré d'implication du locuteur ? La notion de « position discursive » a été définie précisément par Berman en 2005. Elle est composée de trois éléments : l'orientation, l'attitude et le degré de généralité. L'orientation précise la relation existant entre l'émetteur, le récepteur et le texte, qui participent tous trois à la production et à l'interprétation du texte. Ainsi, un texte peut être fortement subjectif, en étant orienté vers l'émetteur, avec une implication personnelle importante de ce dernier ; il peut être orienté vers le récepteur, ce qui se remarquera par des adresses directes comme les particules énonciatives « tu sais » ; enfin, le locuteur peut utiliser des termes plus détachés, en prenant pour point de référence non pas un des deux participants de la situation de communication, mais le texte lui-même. Ensuite, le locuteur peut avoir diverses attitudes envers son texte. Celle-ci peut être épistémique, indiquant dans quelle mesure le locuteur considère comme vrai ce qu'il dit, pouvant utiliser des modalisateurs pour nuancer la véracité de ses propos ; l'attitude peut être déontique, exprimant un jugement de la part du locuteur ; enfin, ce dernier peut laisser transparaitre ses émotions, son attitude étant alors plutôt affective.

Ces trois attitudes permettent ainsi de former un continuum entre des réactions fortement subjectives et des propos davantage objectifs. La dernière composante de la position discursive, le degré de généralité, s'intéresse aux divers actants du récit : participants humains, lieux, moments, objets, etc. Leur référence peut être personnelle, spécifique : le récepteur sait alors précisément de qui et de quoi le locuteur est en train de parler. En revanche, l'émetteur peut choisir des termes plus génériques, qui perdent alors toute spécificité. Enfin, certains textes comportent également des pronoms impersonnels, qui sont non référentiels.

2.2. Différences dues au type de texte

Nos corpus sont composés de deux types de textes : un texte expositif, portant sur la thématique des conflits entre les gens, et deux textes narratifs. Le premier texte narratif est un récit d'expérience d'une situation de problème avec d'autres personnes ; le deuxième est la suite d'un récit fictif relatant l'évasion d'un prisonnier, nommé « Dimitri ».

Tout d'abord, nous pouvons indiquer que le cout lié à la planification diffère fortement entre ces deux types de textes. En effet, le texte narratif est un des rares cas où la structure est relativement linéaire dans la représentation mentale, puisque le scripteur peut se baser sur la structure temporelle des évènements pour construire l'enchainement des propositions (Fayol, 1997). Ce n'est pas le cas de la production du texte expositif, pour laquelle il s'agit de linéariser une représentation multidimensionnelle. Cela rejoint les types de planifications distingués par Bereiter et Scardamalia (1987), « stratégie des connaissances rapportées » et « stratégie des connaissances transformées »[9]. De plus, une des difficultés concernant cette planification tient au fait que même si un texte narratif a un contenu qui n'est pas intéressant, il sera compréhensible si la structure du texte est correcte ; ce n'est pas le cas du texte expositif, dont la structure est fortement liée au contenu, celui-ci devant donc forcément être pertinent pour rendre l'ensemble compréhensible (Berman & Verhoeven, 2002).

D'un point de vue syntaxique, il a été noté que les textes narratifs comportaient davantage de clauses juxtaposées et coordonnées que les textes expositifs, ces derniers privilégiant donc les structures subordonnées (Akinci, 2005; Mazur-Palandre, 2007). Par ailleurs, des structures verbales souvent considérées comme plus complexes, telles que la forme passive ou les verbes non-finis, sont également plus fréquentes dans les textes expositifs (Johansson, 2009; van Hell, Verhoeven, Tak & van Oosterhout, 2005).

L'usage des connecteurs diffère entre les deux types de textes. Tout d'abord, suite aux remarques précédentes quant à la succession temporelle des évènements relatés, il apparait évident que les connecteurs temporels sont plus nombreux dans les textes narratifs qu'expositifs (Ramasawmy, 2004;

[9] Cette distinction sera explicitée dans le chapitre 2.

Schneuwly, Rosat & Dolz, 1989), ces connecteurs permettant de souligner la structure du texte narratif (Bestgen & Costermans, 1997). Les connecteurs temporels peuvent indiquer la simultanéité de l'action (« pendant ») ou une rupture dynamique (« soudain ») (Fayol, 1985). La proportion de connecteurs adversatifs est la même dans ces deux types de textes (Ramasawmy, 2004), mais avec un sens légèrement différent : par exemple, dans un récit, le connecteur « mais » est utilisé pour indiquer la présence d'un obstacle statique (Mouchon, Fayol & Gombert, 1989). Par ailleurs, la présence de connecteurs est indispensable pour assurer la cohérence d'un texte expositif, car ils permettent d'expliciter les relations entre les propositions. Ce n'est pas forcément le cas pour un texte narratif, si la succession des énoncés reflète la succession temporelle des évènements ; d'ailleurs, le nombre de connecteurs dans les récits diminue avec l'âge (Favart & Chanquoy, 2007; Schneuwly, 1988).

Enfin, certaines différences dues au type de texte se reflètent par des choix lexicaux. Par exemple, l'orientation vers le locuteur est plus grande dans le cas de récits d'expérience personnelle que dans des textes expositifs ; il y a donc davantage de pronoms personnels dans les textes narratifs que dans les textes expositifs, où la production est plus distanciée (Berman, 2004b). En outre, le degré de généralité supérieur des textes expositifs implique davantage d'informations génériques et de noms abstraits dans ce type de texte (Berman & Nir-Sagiv, 2004). Par ailleurs, l'ancrage énonciatif spécifique à chacun de ces types de textes implique que les temps verbaux utilisés diffèrent. Ainsi, un récit relatant une expérience, même si elle est au passée, est dans un plan davantage embrayé qu'un récit fictif (Touratier, 1989) ; les textes expositifs se distinguent également par un plan non embrayé, et donc par exemple par l'usage de présents exprimant des vérités générales, et non des présents associés à la situation d'énonciation. Ces différences en production sont visibles même chez les scripteurs les plus jeunes, mais seuls les adultes sont en mesure d'utiliser les différences de temps verbaux pour effectuer une tâche de catégorisation de textes (Benoit & Fayol, 1989).

2.3. Différences dues à l'âge

Un de nos corpus présente des données développementales. Notre intérêt porte non pas sur l'acquisition du langage, mais sur le « développement tardif du langage », c'est-à-dire la manière dont les compétences langagières continuent à se développer chez les enfants et les adolescents. En effet, au cours de l'école primaire, les textes présentent une augmentation du nombre de mots et d'idées, ainsi qu'une complexification syntaxique (Puranik, Lombardino & Altmann, 2008). Ensuite, le développement se poursuit grâce à l'automatisation de certains processus rédactionnels, l'augmentation des capacités cognitives, un plus grand contact avec la modalité écrite du langage (Berman, 2004b) mais également grâce à des activités sociales et à l'enseignement dispensé (Nippold, 2004). À 17 ans, les compétences linguistiques des adolescents se rapprochent de celles des adultes, mais sans qu'elles leur soient encore équivalentes (Johansson, 2009) ; des études portant sur l'acquisition d'une expertise professionnelle présentent ainsi un pattern développemental au sein de groupes d'adultes (Johansson,

Gustafsson & Wengelin, 2015). Pour Berman (2008), être « locuteur expert » suppose que les formes linguistiques disponibles dans le répertoire lexico-grammatical soient utilisées de manière flexible et appropriée dans des contextes de communication variés. Cela ne peut se faire sans une expérience pragmatique et linguistique assez riche. Ce développement de la langue se remarque notamment par la manière dont le texte est construit, le contenu référentiel, les connecteurs entre clauses, les items linguistiques choisis ainsi que la position discursive adoptée. Ces aspects ont particulièrement été analysés dans le cadre du projet Spencer[10], dont de nombreuses études permettent de refléter le développement de plusieurs aspects linguistiques (pour la revue d'un certain nombre d'éléments, voir Berman, 2008).

Nous avons évoqué dans la section précédente les différences dues au type de texte. Nous pouvons à présent mentionner la manière dont se développe justement cette adéquation entre caractéristiques linguistiques et type de texte. Tout d'abord, au niveau de l'identification, il a été remarqué que les enfants de 9 ans distinguaient les textes des non-textes, puis qu'à 11 ans ils parvenaient à isoler les textes narratifs, mais qu'il fallait attendre 13 ans pour qu'ils isolent correctement les textes argumentatifs (Benoit & Fayol, 1989)[11]. Au niveau de la production, les scripteurs sont capables assez jeunes de produire des récits (Berman & Slobin, 1994; Jisa, 2000), même si leur production évolue durant plusieurs années pour passer de la juxtaposition d'évènements indépendants au respect du schéma textuel canonique (Fayol, 1987) ; le texte expositif est en revanche maitrisé plus tardivement (Berman & Katzenberger, 2004). Ces différences développementales peuvent tout d'abord provenir de la connaissance du genre, celle-ci étant liée à l'enseignement (Donovan & Smolkin, 2006) et à la pratique de ces types de texte. Or le récit est un genre assez familier pour l'enfant (contes, récits d'évènements, anecdotes quotidiennes), qui en possède donc un certain nombre de connaissances discursives. À l'inverse, le texte expositif est peu présent dans la vie quotidienne mais davantage proposé à l'école, ce qui explique d'ailleurs qu'il soit davantage associé à un langage académique ou à des caractéristiques de l'écrit, modalité avec laquelle les élèves découvrent principalement ce type de texte. Ainsi, alors que la macrostructure narrative est apprise assez tôt, des éléments constitutifs du schéma textuel expositif, comme la conclusion, sont maitrisés plus tardivement (Katzenberger, 2005) [12]. De plus, cette différence peut être due aux capacités cognitives des enfants et adolescents, qui permettent la généralisation puis l'abstraction des propos. En effet, il y a à l'adolescence une croissance et un changement significatifs

[10] Il s'agit d'un projet financé par la fondation Spencer entre 1997 et 2001, intitulé « Developing literacy in difference contexts and different languages », porté par Ruth Berman. Les productions consistent en des textes écrits et oraux, narratifs et expositifs, écrits par des adolescents de 10 à 17 ans, et produits en anglais, français, espagnol, néerlandais, suédois, hébreu ou islandais.

[11] Les programmes d'enseignement proposés par le Bulletin Officiel spécial n°11 du 26 novembre 2015 (désormais « BO de 2015 ») indique que dès le cycle 2 (CP – CE2), les élèves doivent pouvoir assurer un « repérage et une prise en compte des différents genres de discours (récit, compte-rendu, reformulation, exposé, argumentation » lors de productions orales.

[12] Les nouveaux programmes proposés par le BO de 2015 détaillent les composantes de l'argumentation au cycle 4 (5ème – 3ème), cycle pendant lequel sont également abordés les « indicateurs de temps, de lieu, de relations logiques ».

de certaines régions du cerveau, comme le cortex pré-frontal, particulièrement lié aux capacités de la mémoire à court-terme, de l'attention et de la planification verbale et comportementale. Cela permet d'expliquer les améliorations des jeunes adolescents dans le domaine du raisonnement (notamment le raisonnement déductif), du traitement de l'information et de l'expertise : il émerge alors davantage de pensées abstraites et planifiées (Steinberg, 2005). Le développement des compétences d'abstractions dans le système cognitif de l'adolescent se retrouve au niveau du contenu thématique : ainsi, il est possible d'observer un passage de propos très concrets, liés à l'école, à des thématiques plus abstraites et générales. Cette capacité d'abstraction permet donc de produire plus facilement des textes expositifs. Enfin, alors que les scripteurs les plus jeunes associent les caractéristiques d'une position discursive à un genre, les plus âgés vont davantage alterner leur position discursive au sein d'un même texte : ils seront ainsi capables d'utiliser des exemples personnels concrets au service de l'argumentation d'un texte expositif, ou d'exprimer des généralités lors d'un discours narratif relatant une expérience personnelle.

D'un point de vue de l'empaquetage syntaxique, Verhoven et van Hell (2008) ont montré que les adultes utilisaient davantage de clauses subordonnées et enchâssées, et Akinci (2005) a mis en évidence que les lycéens et étudiants utilisaient davantage de formes complexes que les élèves de primaire ou de collège, et que le nombre de coordonnées diminuait avec l'âge. Des résultats similaires ont été obtenus par Mazur-Palandre (2007) : une diminution avec l'âge de l'utilisation des juxtaposées et des coordonnées sans ellipse, mais une augmentation des subordonnées et des coordonnées elliptiques. Ce pattern développemental peut être dû au contact plus fréquent avec l'écrit, puisque certaines construction complexes comme les subordonnées non-finies (Gayraud, Jisa & Viguie, 2001), les ellipses du sujet lorsqu'il y a une co-référence (Mazur-Palandre, 2007) et les relatives enchâssées (Gayraud & Martinie, 2004) sont davantage présentes en français à l'écrit qu'à l'oral. La voix passive augmente elle aussi avec l'âge (Jisa et al., 2002). Des constructions également plus complexes apparaissent, par exemple pour éviter de mettre systématiquement le sujet en première position dans la phrase : celle-ci peut débuter par un complément circonstanciel, par un participe passé permettant de caractériser le sujet, ou le scripteur peut choisir de faire une inversion du verbe et du sujet (Myhill, 2009).

L'utilisation des connecteurs est caractérisée tout d'abord par une diversification de ces derniers : « et » est utilisé très tôt, mais peut être considéré comme un « archiconnecteur » tant ses usages sémantiques sont divers ; puis le locuteur utilise les connecteurs temporels, puis causaux, l'adversatif « mais », puis les restrictifs (Favart & Passerault, 1999). Cette diversification témoigne du passage d'une organisation linéaire à une structure davantage hiérarchique, nécessaire notamment pour les textes argumentatifs ou expositifs. D'ailleurs, l'étude d'Akiguet et Piolat (1996) indique que les connecteurs adversatifs et conclusifs sont correctement utilisés par les enfants de 11 ans, mais pas de 9 et 10 ans. En outre, certaines relations conceptuelles sont plus difficiles à acquérir que d'autres, ce qui peut ensuite être reflété par les usages des connecteurs : ainsi, les relations causales négatives sont maitrisées plus

tardivement que les relations causales positives ou que les relations additives négatives (Spooren & Sanders, 2008). Cependant, le développement ne se caractérise pas par une augmentation du nombre de connecteurs utilisés : au contraire, Schneuwly et al. (1989) notent une diminution du nombre d'organisateurs textuels entre 10, 12 et 14 ans.

Même si les enfants sont capables de distinguer assez tôt s'il y a ou non une continuité référentielle dans une production, la capacité à introduire, maintenir et réintroduire des référents se développe graduellement (Jisa, 2004). Par exemple, il est nécessaire de prendre en compte l'accessibilité du référent du point de vue du destinataire pour utiliser le bon item, or cette prise en compte du destinataire ne serait effective qu'à partir de 9 ans et continuerait d'évoluer par la suite (Favart & Passerault, 1996)[13], notamment par une diversification des marques de référence. Ainsi, si les plus jeunes privilégient le maintien du référent grâce à des formes pronominales, ils utilisent ensuite des ellipses dans des structures avec un verbe fini puis des clauses non finies (Jisa, 2004) ; en outre, ils parviennent par la suite à maintenir les référents davantage sous forme lexicale, notamment en les diversifiant (Mazur-Palandre, 2009)[14]. De manière générale, les outils linguistiques permettant d'assurer la cohérence et la cohésion du texte sont particulièrement pertinents pour étudier le développement tardif du langage (Tolchinsky, 2004). Cependant, les résultats peuvent dans un premier temps apparaitre surprenants. Nous avions en effet évoqué l'absence de lien clair entre quantité d'outils cohésifs et qualité des productions : McNamara et al. (2010) notent ainsi que les lycéens les plus jeunes ont les productions les plus cohésives, mais de moins bonne qualité, ce qui montre que la maitrise des outils de cohésion continue à se développer même chez les adolescents les plus âgés. Il convient alors davantage d'examiner ces éléments de manière qualitative, par exemple en évaluant leur diversification et leur utilisation correcte, qu'en termes quantitatifs.

Enfin, quelques aspects sémantiques et lexicaux, reflétant notamment la position discursive du locuteur, évoluent avec l'âge. Cela est notamment dû aux changements cognitifs évoqués précédemment. Par exemple, on passe de termes lexicaux plutôt liés au vocabulaire quotidien de l'école à des notions plus abstraites (Ravid & Cahana-Amitay, 2005). Par ailleurs, la densité lexicale des clauses, c'est-à-dire la proportion d'items lexicaux, augmente avec l'âge (Mazur-Palandre, 2009).

[13] Le BO de 2015 insiste sur la prise en compte du destinataire essentiellement lors du cycle 4 (5ème – 3ème)

[14] Le BO de 2015 indique qu'en cycle 4 (5ème – 3ème) doivent être abordés l' « identification et l'utilisation des éléments linguistiques de cohésion textuelle : substituts nominaux et pronoms de reprise, procédés de désignation et de caractérisation, rôle des déterminants indéfinis et définis »

3. Les unités de segmentation du texte

Le scripteur part donc d'une unité que nous avons identifiée comme étant la clause, dont le noyau est le prédicat. Cependant, si le but est de produire un texte, il est nécessaire que l'ensemble des propositions forme un tout cohérent, c'est-à-dire lié d'un point de vue sémantique, syntaxique, et répondant aux normes du schéma textuel attendu pour le type de texte produit. La deuxième partie de ce chapitre nous a montré l'étendue des éléments à prendre en compte de la part du scripteur pour produire son texte, et de la part du chercheur pour rendre compte des compétences textuelles du scripteur. On peut alors chercher à segmenter le texte produit : cette segmentation peut permettre d'une part d'identifier les unités proposées au destinataire, qui pourront servir de base lors du traitement de compréhension, et d'autre part d'identifier les unités de production du scripteur, c'est-à-dire l'étendue du texte sur laquelle le scripteur a réfléchi avant de la produire. Or, Le Goffic (2006, p. 90) résume parfaitement les difficultés auxquelles nous sommes confrontés face à cette question : « La recherche des unités d'un texte est à la fois une nécessité évidente (comment traiter un texte si on ne le découpe pas ?) et une source de contradictions inévitables : qui dit unité dit individualisation, indépendance, autonomie, mais qui dit texte (tissu) dit interdépendance des parties. » Nous allons donc tout d'abord présenter quelques unités qui ont été identifiées pour permettre l'étude des textes ; cela permettra notamment d'affiner notre terminologie. Nous montrerons ensuite quelles sont les difficultés auxquelles est confronté le scripteur quand il doit lui-même segmenter son texte, lors de sa production.

3.1. Les unités identifiées lors de l'étude d'un texte

Si l'on part des plus petites unités, à savoir les clauses, il est possible de délimiter des unités supérieures uniquement définies en termes syntaxiques. Ainsi, Hunt (1970, p. 4) a introduit une unité qu'il nomme « minimal terminable unit » ou « T-unit », et qui a une définition uniquement syntaxique : une « Unité Terminale » (ou unité T) englobe des clauses entretenant des relations de dépendance, donc les clauses principales avec leur(s) subordonnée(s). Il est cependant possible d'étendre cette notion de « dépendance », en considérant qu'une clause coordonnée contenant une ellipse du sujet est dépendante de la clause précédente (Mazur-Palandre 2009). Cette notion d'unité définie en termes syntaxiques correspond à ce que Paolacci et Rossi-Gensane (2014; 2012) désignent comme « phrase syntaxique ».

Ces phrases syntaxiques peuvent parfois être coordonnées entre elles ou simplement juxtaposées, formant des unités supérieures typographiquement ou prosodiquement démarquées. Nous rejoignons ainsi la notion de « période », notamment utilisée dans le cadre d'études sur l'oral, où le critère typographique ne fonctionne pas. Cette notion de « période » regroupe des clauses selon un critère qui, pour Adam (1991), est essentiellement rythmique ; les unités composant la période sont liées soit par le sens, soit par l'utilisation de connecteurs. Généralement, la fin d'une période est donc marquée soit par l'intonation (à l'oral), soit par une marque de ponctuation forte (à l'écrit). Le fait de recourir ainsi à un

découpage graphique ou intonatif permet alors au locuteur de produire des « effets de sens » (Adam, 2011a) pour le destinataire. Cette notion de « période » est proche de ce qui est défini par le modèle de Fribourg, qui se base davantage sur une approche pragmatique, mais qui indique que la période, signalée à l'oral par un intonème conclusif, réalise un programme énonciatif complet (Berrendonner, 2002).

Dans un texte, les périodes sont regroupées au sein de macropropositions (e.g. la complication ou la résolution dans un texte narratif). Ces grandes unités sont sans doute celles identifiées par le scripteur lorsqu'il doit imaginer et organiser l'essentiel de son texte ; c'est également ce qui doit être retenu par le destinataire lorsqu'il reconstruit sa représentation mentale. Ces macropropositions sont constitutives du schéma textuel, Adam (1993) identifiant des « séquences » caractérisées par un schéma prototypique (e.g. séquence narrative, séquence descriptive).

En s'inspirant de Adam (1993), nous pouvons donc identifier la structure hiérarchique suivante : [texte [séquence [macroproposition [période [unité Terminale [clause]]]]]]. La segmentation de ces unités est particulièrement difficile au sein de la macroproposition. Dans notre chapitre « Méthodologie générale », nous expliciterons ainsi quels critères nous ont permis de segmenter les textes en clauses, puis comment nous avons distingué deux niveaux à partir de critères syntaxiques : l'unité Autonome, comprenant une clause et, le cas échéant, les clauses qui lui sont subordonnées, et l'unité Terminale, comprenant une unité Autonome et, le cas échéant, les unités Autonomes adjacentes comprenant une ellipse du sujet. En outre, il est particulièrement difficile de définir ce qu'est une « période » ; cette notion est souvent liée à celle de « phrase ». Nous allons donc à présent montrer dans quelle mesure cette notion de « phrase » est difficile à définir, enseigner, et par conséquent les problèmes de segmentation éprouvés par les scripteurs.

3.2. Les difficultés de segmentation lors de l'écriture d'un texte

3.2.1. La « phrase », une unité difficile à définir

Lorsque l'on essaie de définir ce qu'est une phrase, de multiples critères peuvent être évoqués : réalisation d'un acte de langage, complétude sémantique ou syntaxique, construction interne, démarcation intonative ou typographique ; le problème est qu'aucun de ces critères ne semble ni nécessaire ni suffisant (Barjolle, 2001; Berrendonner, 1993). Une des raisons de cette difficulté est que la phrase est à la frontière des domaines syntaxique et textuel (Barjolle, 2001; Le Goffic, 2001), et un des écueils, notamment au niveau didactique, est alors de ne considérer que les phrases « en soi » et non au sein d'un texte (Rondelli, 2013). Pourtant, la phrase semble être une notion de sens commun (Béguelin, 2002) ; il serait alors tentant de la décrire en termes de « prototypes » pour rendre compte de cette unité qui semble « réelle » pour les sujets ; cela est cependant peu convaincant dans un cadre scientifique (Kleiber, 2003).

Il apparait alors que, dans beaucoup de cas, c'est le critère prosodique ou typographique qui permet finalement d'affirmer ce que le locuteur a souhaité considérer comme « phrase ». En effet, même pour des élèves qui n'arrivent pas à se mettre d'accord entre eux sur ce qu'est une phrase lors de la production d'un texte, elle apparait toujours de façon claire a posteriori, comme un segment situé entre deux points (Boré & Bosredon, 2013). Outre les différents critères mentionnés ci-dessus, le choix prosodique ou typographique peut être lié à des motivations « esthétiques » telle que la longueur des segments, ou par la volonté de mettre en relief de manière perceptive certains éléments, en rendant par exemple autonome typographiquement ou prosodiquement un constituant syntaxique qui est normalement lié aux autres constituants (Béguelin, 2002). En outre, la prise en compte du destinataire peut amener le scripteur à choisir comme « phrase » un segment qui pourra être traité en mémoire comme une unité (Berrendonner, 1993). Ainsi, comme notre travail porte sur des productions écrites, nous parlerons désormais de « phrase » dans le sens de « phrase graphique » (débutant par une majuscule et se terminant par un point). Certes, cette définition n'est pas parfaite, puisqu'il y a une définition circulaire entre le fait que la « phrase finit par un point » et qu'un « point sert à finir une phrase » (Béguelin, 2002) ; cependant, dans le cadre de notre étude la « phrase (graphique) » n'est qu'une unité parmi d'autres. Il n'y a d'ailleurs pas d'équivalent direct entre phrase graphique et période (la période pouvant être plus longue que la phrase graphique) ni entre phrase graphique et clause (Berrendonner, 1993). D'un point de vue développemental, les unités Terminales ont tendance à augmenter avec l'âge (Hunt, 1965, 1970) alors que le nombre d'unités T par phrase diminue. Cela rejoint le constat de Paolacci et Rossi-Gensane (2014; 2012) qui relèvent de nombreux cas de sous-ponctuation, les phrases graphiques étant beaucoup plus longues que les phrases syntaxiques[15], et qui indiquent que les textes de meilleure qualité sont ceux où le nombre de phrases syntaxiques se rapproche du nombre de phrases graphiques.

La question de la maitrise de la phrase d'un point de vue développemental est liée à l'enseignement qu'il en est fait. La notion de « phrase » est notamment étroitement liée aux critères typographiques[16]. Or, les élèves considèrent que le respect de la norme typographique est très important pour leur enseignant, alors que ces derniers soulignent certes l'importance d'achever les phrases par des signes de ponctuation forts, mais prennent peu en compte ces erreurs lors de leurs évaluations (Bosredon, 2014). En outre, l'importance de la ponctuation « normée » laisse peu de place à l'enseignement des usages flottants, par exemple par la présentation de styles d'écritures différents ou d'écrivains professionnels ne respectant justement pas ces normes (Dufour, 2014). Dès lors, on voit que la notion de phrase est confuse pour les élèves : ainsi, ceux de CM2 vont davantage s'attarder sur des aspects quantitatifs ou sur la longueur des phrases que sur une réflexion portant sur la complétude sémantique ou syntaxique,

[15] équivalents de l'unité Terminale définie par Hunt

[16] Le Bulletin Officiel hors-série n°3 du 19 juin 2008 suggère qu'en CP les élèves sachent « identifier les phrases d'un texte en s'appuyant sur la ponctuation (point et majuscule) ». Le BO de 2015 indique simplement que les élèves de cycle 2 doivent être en mesure d'identifier la « ponctuation de fin de phrases » et les « signes du discours rapporté ».

ou sur l'effet que peut avoir la ponctuation sur le lecteur (Colognesi & Deschepper, 2014). Même pour les lycéens, la notion de « phrase complexe » est associée à une phrase longue et difficile, le critère quantitatif semblant donc prédominer (Chanfrault-Duchet, 2001).

3.2.2. Le rôle de la ponctuation dans la connexion-segmentation du texte

La prise en compte de la ponctuation pour définir les phrases nécessite de s'attarder sur le rôle du système de ponctuation pour assurer la connexion du texte. En effet, le lien entre les clauses est indiqué non seulement par les connecteurs, mais également par les signes de ponctuation, qui permettent d'indiquer le degré de liaison entre les clauses, et également de regrouper syntaxiquement plusieurs clauses (Coirier et al., 1996). La ponctuation peut être vue comme marquant une rupture dans le texte, cette rupture étant plus ou moins forte en fonction de la place dans la hiérarchie : alinéa > point > point-virgule > virgule > absence de signe (Favart & Chanquoy, 2007).

Cette hiérarchie n'est pas maitrisée par les scripteurs les plus jeunes. La ponctuation sert tout d'abord à délimiter des propositions, puis des groupes de propositions (Passerault, 1991). Même si la proportion de ponctuation est stable entre le CP et le CE2, on observe une utilisation plus normée de ces marques en CE2 (Chanquoy & Fayol, 1995). Cependant, la virgule apparait bien comme un « archiponctème » (Dufour, 2014) puisqu'elle est utilisée très fréquemment, les scripteurs de 8-9 ans ne saisissant pas la distinction de degré de rupture entre la virgule et le point (Favart & Chanquoy, 2007). Les autres signes de ponctuation sont peu utilisés même en CM2, et un étayage lié à ces différents signes peut permettre une diversification de la ponctuation, mais avec un usage parfois fautif (Dufour, 2014). À la fin de l'école primaire, les élèves semblent connaitre les règles, mais ils ont du mal à les mettre en pratique : les majuscules et les points délimitant les phrases ne sont pas toujours correctement indiqués (Fauchart, 2001), la ponctuation gérant le discours direct (guillemets, deux-points) n'est pas toujours correctement disposée (Passerault, 1991), et les alinéas sont parfois utilisés de manière massive et non justifiée (Paolacci & Favart, 2010). Les programmes scolaires du BO de 2015 indiquent que l'« analyse du rôle syntaxique des signes de ponctuation et l'utilisation de ces signes à bon escient » est un objectif du cycle 4 ($5^{ème}$ – $3^{ème}$).

Une difficulté quant à la bonne gestion de la segmentation vient du fait que la ponctuation n'est pas qu'une indication de degré de connexion. Par exemple, certaines virgules sont indispensables d'un point de vue sémantique pour distinguer les relatives restrictives des relatives explicatives. En outre, si la ponctuation forte peut avoir une valeur prosodique équivalant à ce qui serait une pause longue à l'oral, le passage de l'oral à l'écrit n'est pourtant pas biunivoque (Béguelin, 2002), certaines pauses longues ne correspondant pas à des points, et vice versa.

3.2.3. La phrase : une unité de réception ou de production ?

Ainsi, il semble relativement aisé d'identifier les phrases en fonction de critères typographiques. Le rôle de la segmentation pour le lecteur a plusieurs fois été mis en avant : la ponctuation sert à détacher des unités pour indiquer la structure du texte, ce qui n'est d'ailleurs pas toujours facile puisqu'il est alors nécessaire d'anticiper les besoins du lecteur, le scripteur devant devenir son propre lecteur pour effectuer une segmentation correcte (Jaffré, 2014; Paolacci & Rossi-Gensane, 2012).

Cependant, si la phrase est une unité facilement identifiable pour le destinataire lors du décodage, elle est beaucoup plus instable lors de l'encodage, le scripteur pouvant éprouver des difficultés à poser des choix liés à la segmentation de son texte (Colognesi & Deschepper, 2014; Paolacci & Rossi-Gensane, 2012 ; 2014). Or, Dufour (2014) relève que les scripteurs ponctuent leur texte essentiellement pendant l'écriture, et non au cours d'une relecture ultérieure. Passerault (1991, p. 91), quant à lui, propose que « la ponctuation constitue le lieu privilégié d'accomplissement de l'activité de planification », ce qui rejoint la proposition de Rondelli (2013) selon laquelle les phrases constituent des unités de traitement pour le lecteur, mais qui peuvent être posées comme d'éventuelles unités de production pour les scripteurs. La question se pose alors du lien entre les unités de segmentation telles qu'elles sont identifiables sur le produit, et les unités de production.

4. Conclusion

Ce chapitre a permis de montrer ce qu'était un « texte », et notamment dans quelle mesure un texte, de par sa cohérence, se distinguait d'une simple juxtaposition de clauses. Les contraintes linguistiques et discursives pesant sur le scripteur rendent l'activité de production écrite difficile. En effet, le scripteur doit former des clauses grammaticalement correctes, les lier d'un point de vue sémantique et syntaxique, et enfin se conformer au schéma textuel attendu. Pour cela, se posent à chaque instant des choix sémantiques, syntaxiques et lexicaux, qui dépendent également du type de texte et de l'âge des scripteurs. Les caractéristiques linguistiques évoquées dans les parties précédentes ne constituent pas la liste exhaustive de ce qui peut être étudié en psycholinguistique textuelle. Cependant, elles reflètent la complexité de ce qu'est un texte. Un certain nombre de ces éléments seront utilisés dans nos analyses pour comprendre dans quelle mesure certaines unités sont plus ou moins difficiles à produire. Par ailleurs, il arrivera que des éléments ne soient pas étudiés directement, mais ils pourront aider à interpréter certains résultats. On peut en effet faire l'hypothèse que certains de ces choix sont plus ou moins faciles : par exemple, il pourrait être plus facile de choisir des mots fréquents, mais plus difficile d'utiliser la forme passive du verbe. Cette problématique de la gestion de la production en fonction des caractéristiques du texte sera donc abordée dans notre travail. Par ailleurs, la présentation des unités de segmentation du texte a commencé à faire émerger une autre question, essentielle à notre travail : celle des unités de production. Nous avons en effet vu que certaines unités étaient relativement faciles à identifier du point de vue du lecteur, une fois les critères clairement définis ; cependant, l'hypothèse selon laquelle les unités de segmentation ainsi définies (clauses, phrases) correspondent à des unités de production reste à explorer.

Chapitre 2 - Produire un texte : la complexité des processus rédactionnels

Le scripteur cherchant à produire un texte a donc pour but de construire un ensemble répondant à un certain nombre de contraintes linguistiques, telles que présentées précédemment. Cependant, l'analyse du produit fini peut masquer le fait que celui-ci est l'aboutissement d'une activité complexe, qui peut être appréhendée par l'étude des processus rédactionnels.

La recherche sur les processus rédactionnels nécessite un cadre permettant d'identifier les diverses composantes de la production écrite, ainsi que le recours à différentes techniques. Nous présenterons tout d'abord les modèles dont nous disposons pour rendre compte de la complexité de ce type de tâche : cela nous permettra d'identifier les processus rédactionnels auxquels nous nous intéresserons, et d'introduire la problématique de la gestion de ces processus, notamment leur gestion temporelle. Plane (2004) a émis des réserves quant à l'usage de modélisations, du fait qu'elles ne pouvaient pas rendre compte de l'ensemble des multiples dimensions de l'écriture, et que le risque était d'influencer trop fortement les interprétations. Cependant, elle souligne que « si on ne perd pas de vue le fait que les modélisations sont par nature sommaires, et que c'est en cela qu'elles sont utiles, parce qu'elles constituent des schématisations rendant visibles des points noyés dans la complexité du réel, on peut en faire un bon usage » (p. 80). C'est dans cet esprit que nous présenterons les cadres conceptuels qui permettront d'identifier les composantes de la production écrite et de les étudier. Les modèles présentés, bien que multiples et différant parfois quant aux terminologies employées, sont le plus souvent complémentaires (pour une revue de certains modèles, voir Alamargot & Chanquoy, 2001, 2002; Coirier et al., 1996; Fayol, 1997; Fayol & Heurley, 1995; Ferrand, 2002; Leuwers, 2002).

1. Le modèle de Hayes et Flower et son évolution

Un premier modèle, largement utilisé par une grande partie de la communauté scientifique depuis sa publication, est celui de Hayes et Flower (1980), également décrit en détail dans l'article de Flower et Hayes (1981) ; la Figure 2 le présente dans sa version française[17]. Ce modèle donne un cadre conceptuel en identifiant les processus intervenant lors de la production écrite d'un texte et leurs relations, ouvrant d'une part la voie à de nombreuses recherches en psychologie cognitive, et permettant d'autre part aux enseignants de prendre conscience des différentes composantes de la production écrite. Ce modèle, élaboré à partir de l'analyse du protocole verbal d'un scripteur, propose que l'écriture implique trois éléments : l'environnement de la tâche (les consignes de la rédaction et le texte déjà écrit), la mémoire à long terme (désormais MLT) du scripteur (comprenant les connaissances liées au thème, au futur lecteur et au type de texte) et les processus rédactionnels. Trois processus rédactionnels sont identifiés : la planification (planning), la formulation (translating) et la révision (reviewing)[18] ; le passage d'un processus à l'autre est géré par le moniteur.

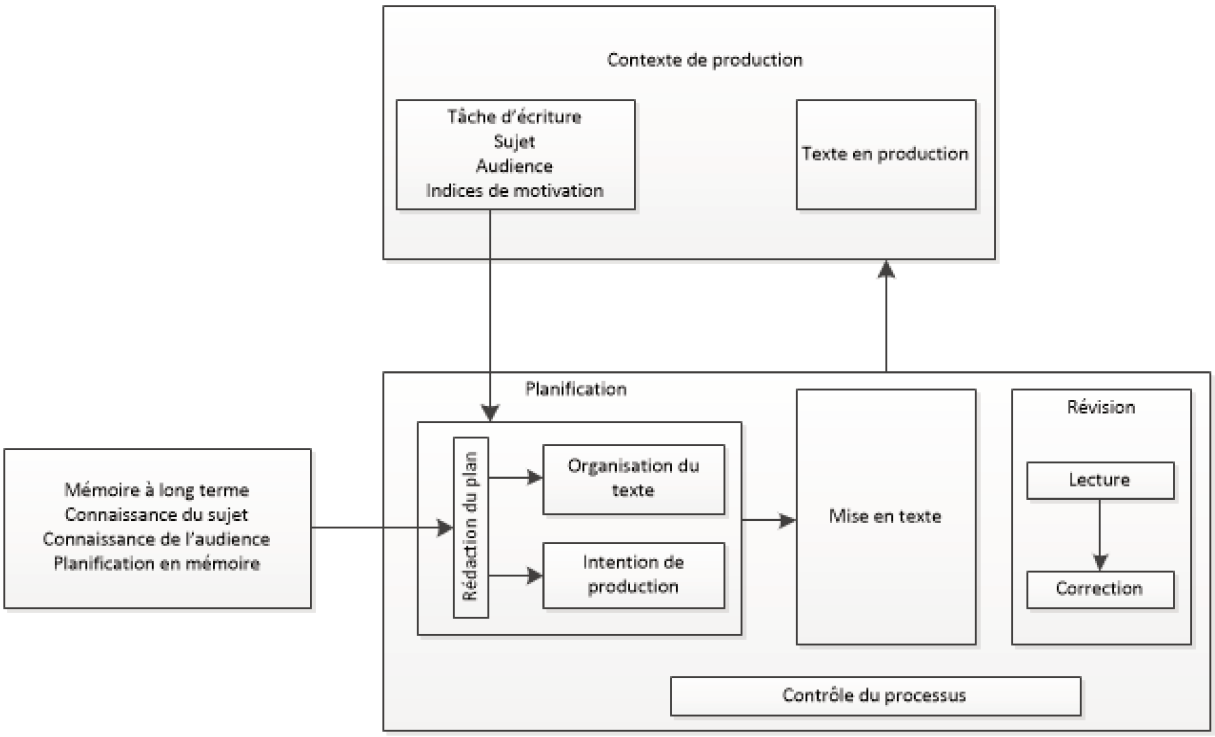


Figure 2. Modèle de Hayes et Flower (1980)

[17] Il s'agit de la traduction proposée par Hayes dans son article de 1995.

[18] Les processus rédactionnels seront décrits en détail dans la section suivante. Les termes anglais sont indiqués car ils peuvent varier d'un auteur à l'autre, avec parfois des nuances, et les traductions françaises ne sont pas toujours identiques. Ainsi, translating est parfois traduit par « mise en texte ».

Plusieurs critiques ont été adressées à ce modèle (Alamargot & Chanquoy, 2001), considéré comme descriptif quant aux composantes de l'écriture, mais peu opérationnel pour expliquer comment, en temps réel, les différents processus sont coordonnés. Cependant, son utilisation par les chercheurs intéressés par la question de la production écrite a justement permis d'en compléter certains aspects. Ainsi, nous verrons par la suite dans quelle mesure l'aspect développemental (Bereiter & Scardamalia, 1987; Berninger & Swanson, 1994), le lien entre les processus rédactionnels et la mémoire de travail (Kellogg, 1996; Kellogg, Turner, Whiteford & Mertens, 2016) ou encore la coordination des processus rédactionnels (Olive, 2014) ont été, et sont encore, sources de recherches. Grâce à ce modèle, la production écrite est envisagée dans son contexte de réalisation, en prenant en compte la singularité de l'individu, et met en avant la complexité des relations entre les différents processus rédactionnels.

La distinction entre le contexte de réalisation et l'individu a été accentuée dans le modèle proposé par Hayes en 1995. Dans le « contexte de production » sont mis en avant le contexte social, précisant plus particulièrement le type de destinataire, et le contexte physique. Dans ce dernier, la mention du « support à l'écriture » introduit la question du medium d'écriture (i.e. ordinateur, feuille), qui n'était pas considérée dans le modèle précédent. Dans la composante « individu » nous retrouvons les connaissances du savoir (liées au sujet, à la langue, aux destinataires) ainsi que les processus cognitifs, ceux-ci ayant été fortement modifiés. En effet, la révision n'est plus considérée comme un processus de base, mais est constituée de plusieurs processus tels que la lecture, la planification et la mise en texte. Le rôle de la planification, quant à lui, a été minoré suite à des études qui auraient pu aboutir à des suggestions pédagogiques, mais qui ont en réalité montré peu de corrélations entre l'attention, les conseils de planification et la qualité des textes produits ; la planification a donc été remplacée par le processus plus large de « résolution de problème ». En outre, ce modèle a pour caractéristique de pouvoir être commun au langage oral et écrit, puisque le processus « traitement du langage » comprend les processus d'écoute et de lecture, et le processus « mise en texte » concerne aussi bien la production écrite qu'orale. Enfin, l'importance de la motivation a été mise en évidence dans des études précédentes, ce qui a conduit Hayes a ajouté cet élément dans la composante « individu ». Ce modèle a pour mérite d'avoir pris en compte les composantes importantes de la production du langage suite aux études menées. Il pose également la question des « processus de base », qui ont été entièrement remodelés. Cependant, on peut lui reprocher d'être finalement peu précis quant aux processus cognitifs, et de ne pas indiquer les relations entre ces processus, hormis pour souligner leur récursivité.

Certaines de ces critiques trouvent leur réponse dans le modèle proposé en 1996, dont une traduction française a été fournie proposée par Hayes en 1998 (Figure 3). Les principales composantes du modèle de 1995 sont présentes : le « contexte de production » (aussi traduit par « environnement de la tâche ») constitué du contexte social et du contexte physique ; l'« individu » constitué de la part motivation / affect, de la MLT et des processus cognitifs. La part « motivation / affect » est davantage détaillée : il

est précisé qu'elle implique le but de la rédaction, les prédispositions du scripteur, ses propres croyances et attitudes et enfin les estimations du cout et du bénéfice de la tâche. Hayes a en outre ajouté à ce modèle la composante « mémoire de travail » (désormais MDT), selon la définition de Baddeley et Hitch (1974). De plus, ce modèle aborde spécifiquement la production écrite, d'où l'importance de la composante « bloc-note visuo-spatial » dans la MDT et le fait qu'il s'agisse d'interprétation et de production du « texte », et non du langage en général.

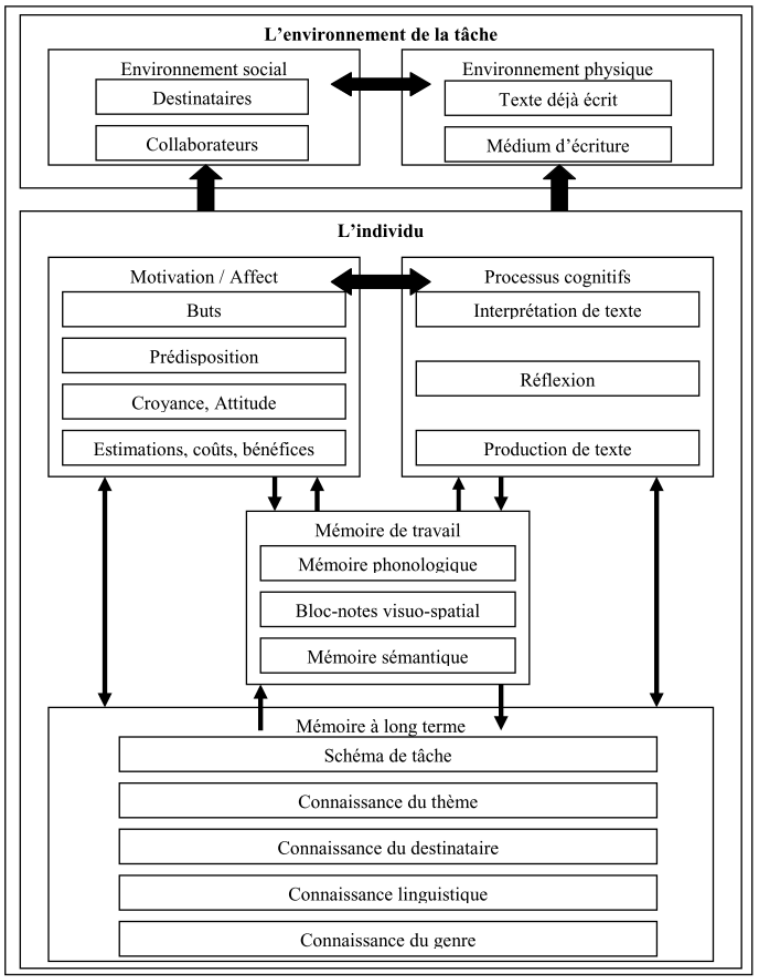


Figure 3. Modèle de Hayes (1996)

Enfin, Chenoweth et Hayes (2001) proposent une évolution notable du modèle, dont une nouvelle version a été publiée récemment (Hayes, 2012). Ce dernier modèle, présenté dans la Figure 4, prend en compte la plupart des remarques faites précédemment : il n'y est plus fait mention des processus de « planification » ou « révision » ; les composantes de la mémoire (MDT et MLT) sont mentionnées ; enfin, le processus de « transcription » est ajouté, suite aux études montrant le cout de ce processus lorsqu'il n'est pas tout à fait automatisé. Ce modèle a pour particularité d'être composé de trois niveaux : niveau de ressources, niveau des processus et niveau de contrôle. Les ressources peuvent être mobilisées par les autres niveaux ; il s'agit de la mémoire (MDT, MLT), de la lecture et de l'attention. Le niveau de contrôle inclut les schémas d'écriture et plan établis, eux-mêmes issus de la production de buts, qui est affectée par la motivation du scripteur. Enfin, le niveau des processus est composé d'une partie

externe, l'« environnement de la tâche », et d'une partie interne, les « processus rédactionnels ». Dans cette dernière composante, nous retrouvons des modules qui correspondent aux opérations mentionnées dès le modèle de 1980 : le proposeur (proposer), le formulateur (translator), l'évaluateur (reviser) et le transcripteur (transcriber).

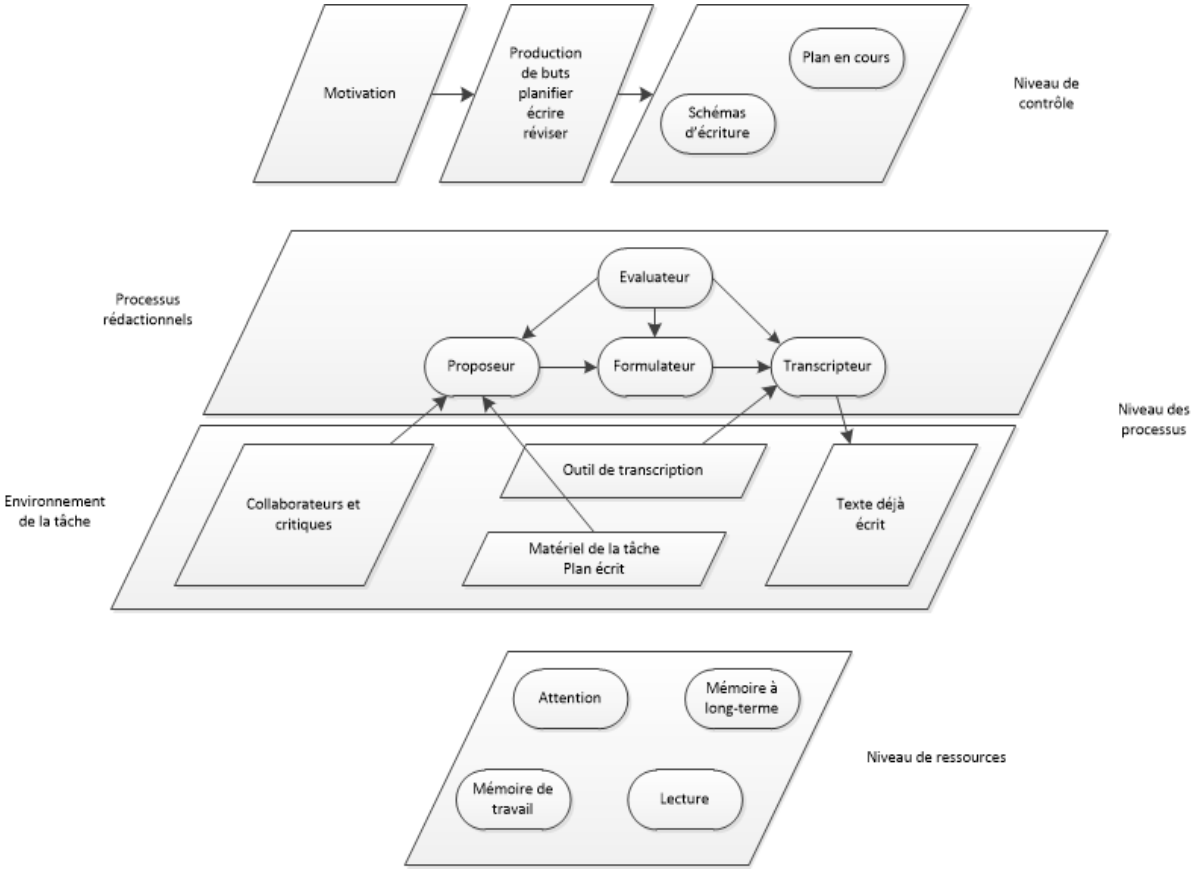


Figure 4. Modèle de Hayes (2012)[19]

Le fait de travailler sur la production écrite nécessite d'avoir conscience de tous les facteurs qui entrent en jeu lors de la rédaction d'un texte : capacités cognitives (de mémoire, d'attention, de lecture), motivation du scripteur, consignes adressées. Ces différents éléments peuvent d'ailleurs expliquer certaines variations dans le contenu du texte produit et / ou les processus mis en œuvre pour cette production. Dans la suite de notre travail, nous nous intéresserons plus particulièrement aux processus rédactionnels eux-mêmes : ce qui avait été appelé initialement « planification », « formulation » et « révision ». Il convient cependant de s'attarder davantage sur leurs définitions. D'une part, ces processus ont fait l'objet de modèles plus précis, qui pourront éclairer nos travaux ; d'autre part, les terminologies diffèrent et ne recouvrent pas toujours les mêmes notions, ce qui rend nécessaire l'explicitation des termes que nous emploierons.

[19] notre traduction

2. L'identification des processus

Parmi les trois processus identifiés dans le modèle initial, qui a servi de cadre à de nombreuses études, nous pouvons distinguer d'une part la planification et la formulation, et d'autre part la révision. Nous choisissons cette répartition car les deux premiers processus sont parfois amenés à se recouper, suivant les modèles et les définitions des auteurs ; la révision, quant à elle, a fait l'objet d'études spécifiques.

2.1. « Planification » et « formulation »

D'après le modèle de Hayes et Flower (1980), le premier processus est la planification : le scripteur a un certain nombre d'idées, qu'il doit organiser. Lors de cette phase, il se sert de différentes sources pour faire émerger les idées pertinentes. Ainsi, dans sa mémoire à long terme, qui constitue un réservoir de connaissances, le scripteur puise parmi celles reliées au sujet de la rédaction, ainsi que parmi les connaissances discursives liées à la production attendue. L'autre source utilisée pendant cette phase de planification est l'environnement de la tâche. Celui-ci permet notamment de définir les enjeux rhétoriques de la production, tels que ses objectifs ou le public visé. Il contient également des sources « extérieures » au scripteur, comme les documents mis à sa disposition ou le texte qui émerge au fur et à mesure de la production. La planification est ainsi composée de plusieurs sous-processus : la génération de contenu et son organisation, et le cadrage des buts qui permet d'ajuster les traitements. Cette notion de « planification » a souvent été associée au fait de « faire un plan », ce qui a conduit aux critiques rapprochant le modèle de Hayes du modèle de la rhétorique classique (Favart & Olive, 2005), et a été utilisé en didactique pour inviter les élèves à « faire un plan » avant d'écrire. Cependant, cette planification ne se fait pas nécessairement sur la globalité du texte, et n'a pas lieu uniquement avant le début de l'écriture.

Une fois les idées organisées, elles doivent être transposées dans la langue choisie : c'est l'étape de formulation. C'est à ce moment-là que le scripteur effectue des choix aux niveaux lexical, sémantique et syntaxique, afin de transmettre à son auditoire ses pensées, qui lors de la phase de planification ne sont pas nécessairement formulées de manière précise, puisqu'il peut s'agir simplement de mots-clés ou d'« images ». Le terme anglais translating insiste sur ce passage d'un message conceptuel pré-linguistique à une forme linguistique. Une des difficultés de cette formulation est la linéarisation, le modèle mental n'ayant pas forcément une organisation linéaire, contrairement à l'énoncé qui sera produit (Fayol, 1989).

Ce processus de formulation est très peu décrit par Hayes, quels que soient ses modèles. On peut cependant avoir recours aux modèles proposés dans le cadre de la production orale, dont le processus de formulation a fait l'objet de plus amples études. Certes, la différence entre ces deux modalités nécessite l'adaptation de certaines composantes. Ainsi, à l'oral apparait un encodage phonétique où

interviennent des processus tels que l'assimilation régressive ou progressive, explicables d'un point de vue articulatoire, alors qu'à l'écrit il y aura un traitement spécifique concernant l'accès à la forme orthographique, puisque le scripteur devra choisir le bon graphème correspondant au phonème à transcrire ; de même, ce qui se joue à l'oral au niveau de l'articulation correspond plutôt, à l'écrit, à la transcription graphique. Plusieurs modèles ont tenté de décrire les processus permettant de passer d'un message pré-verbal à la parole prononcée. De manière sérielle, ils proposent généralement le passage d'une forme sémantique à une structure syntaxique composée de ces éléments sémantiques, puis la sélection des lexèmes appropriés d'un point de vue sémantique et morphosyntaxique avec l'application des règles phonologiques et morpho-phonologiques nécessaires, et enfin la commande motrice permettant d'articuler la forme phonétique extraite des opérations précédentes (Bock & Levelt, 1994; Fromkin, 1971; Levelt, 1989; Roelofs, 1997). Nous présentons dans la Figure 5 le modèle de Levelt de 1999, qui reprend de manière assez détaillée ces différentes étapes.

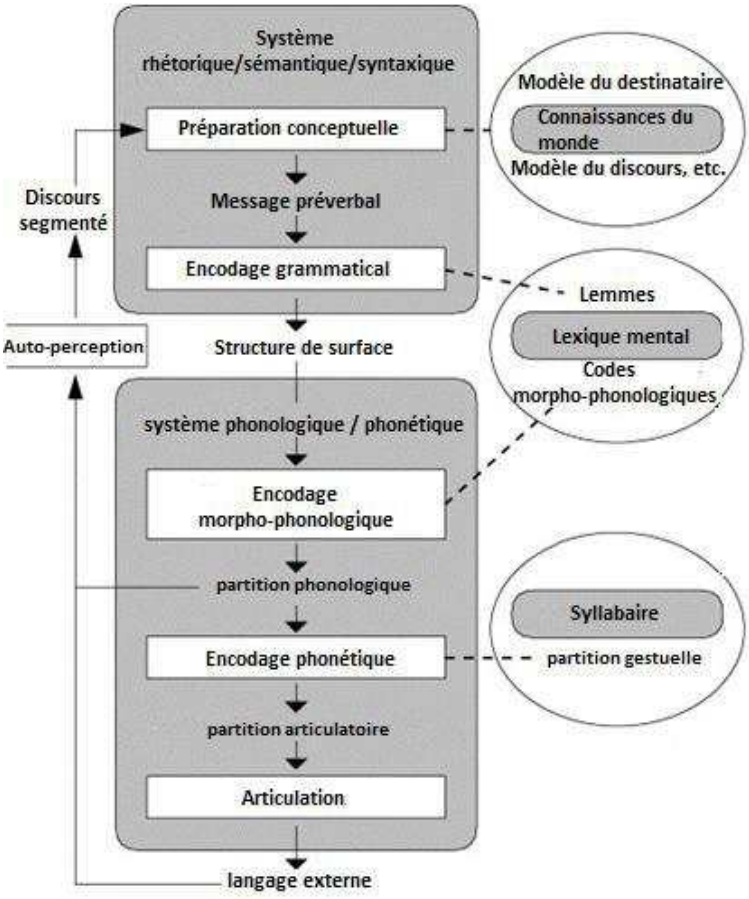


Figure 5. Modèle de Levelt (1999) [20]

Durant la préparation conceptuelle, il y a une première phase de macro-planification, au cours de laquelle le locuteur détermine ce qu'il va dire, à partir de ses connaissances encyclopédiques, discursives ainsi que rhétoriques. Levelt insiste notamment sur le rapport à l'interlocuteur, dans le choix de ce qui

[20] notre traduction

doit être dit et de la perspective adoptée pour faire passer le message souhaité. La micro-planification qui suit permet de traduire les choix adoptés à partir de la situation communicationnelle en choix liés aux relations sémantiques. La syntaxe intervient plus particulièrement au moment de l'encodage grammatical (équivalent de la formulation), qui correspond au passage d'un message préverbal à une structure de surface. En effet, lors de celui-ci, il y a tout d'abord la sélection des lemmes, donc des mots. Ces mots ont des rôles thématiques (agent, patient, expérienceur) auxquels sont attribuées des fonctions syntaxiques : les relations grammaticales ainsi générées reflètent les relations conceptuelles construites dans le message pré-linguistique. Il y a ensuite un encodage morpho-phonologique, suivi de l'encodage phonétique à partir duquel l'articulation se produit.

Les modèles suivant celui de Hayes et Flower (1980) n'ont pas exactement distingué les processus de la même manière. Ainsi, Hayes et Nash (1996) élargissent le concept de « planification » à plusieurs types de traitement, ce qui rejoint la définition de Torrance (2015, p. 1) selon laquelle, de manière fondamentale, « la planification est toute activité cognitive par laquelle une action future est préparée ». Hayes et Nash parlent tout d'abord de la planification des processus, qui concerne la manière d'atteindre les objectifs fixés ; la planification du texte, quant à elle, concerne la forme et le contenu du texte, de manière abstraite, ainsi que la planification linguistique (formes lexicales et syntaxiques utilisées). Dans leur revue des principaux modèles de la production écrite, Alamargot et Chanquoy (2001) proposent quatre étapes pour la partie translating du modèle de Hayes et Flower (1980) : l'élaboration, la linéarisation, la formulation et l'exécution. Les deux premières étapes sont pré-linguistiques ; au niveau de la macro-planification, elles consistent à générer du contenu et à déterminer l'ordre des propositions microstructurales. La formulation « met en mots », dans une structure grammaticale correcte, le message préverbal issu de la linéarisation. Enfin, la dernière étape consiste en l'exécution graphique du produit linguistique. Dans sa proposition d'un modèle en cascade, Olive (2014) distingue la phase de « conceptualisation », qui porte sur les idées ou groupes d'idées, la « formulation » qui convertit l'information sémantique en une structure linguistique, et enfin l'« exécution ». Hayes (1996), lorsqu'il remodèle les processus cognitifs, parle de « réflexion » pour désigner le fait de devoir résoudre un problème, mais également de prendre des décisions concernant son texte. La « production de texte » correspond à la « formulation » (de la récupération sémantique à la mise par écrit de la partie de la phrase concernée).

Ainsi, nous voyons que de nombreux termes sont utilisés pour évoquer les processus rédactionnels. Nous proposons dans le Tableau 1, basé sur le travail de Hayes et Nash (1996) et complété d'après les quelques modèles évoqués ci-dessus, dont la liste n'est pas exhaustive, un aperçu des termes employés. Les frontières entre ces différents termes ne sont pas toujours faciles à déterminer. Dans le cadre de nos recherches, nous distinguerons d'une part la « réflexion » et d'autre part l'« exécution ». La « réflexion » concerne toutes les activités mentales et les prises de décisions nécessaires pour aboutir à

la forme finale. De manière plus précise, nous distinguerons la macro-planification (e.g. planification du scénario, point de vue adopté), la micro-planification (choix sémantiques au niveau local) et la formulation (choix syntaxiques et lexicaux). L'exécution motrice, quant à elle, désigne le fait de mettre par écrit un énoncé.

<table>
<tr><td>Hayes et Nash (1996)</td><td>planification des processus</td><td>planification sans contenu</td><td colspan="2">planification du contenu</td><td>planification linguistique</td><td>production du texte</td></tr>
<tr><td>Hayes et Flower, (1980)</td><td colspan="3">planification</td><td colspan="2">formulation</td><td></td></tr>
<tr><td>Flower et Hayes (1980)</td><td>planification pour composer</td><td>planification pour faire</td><td colspan="2">planification pour dire</td><td colspan="2"></td></tr>
<tr><td>Alamargot et Chanquoy (2001)</td><td></td><td></td><td>élaboration</td><td>linéarisation</td><td>formulation</td><td>exécution</td></tr>
<tr><td>Levelt</td><td></td><td>préparation conceptuelle (macro-planification puis micro-planification)</td><td colspan="2">encodage grammatical</td><td>encodage phonologique, encodage phonétique</td><td>articulation</td></tr>
<tr><td>Chenoweth et Hayes (2001) ; Hayes (2012)</td><td></td><td colspan="3">proposition</td><td>traduction</td><td>exécution</td></tr>
<tr><td>Olive (2014)</td><td></td><td colspan="3">conceptualisation</td><td>formulation</td><td>exécution</td></tr>
<tr><td rowspan="2">Terminologie retenue</td><td rowspan="2"></td><td colspan="4">réflexion</td><td rowspan="2">exécution</td></tr>
<tr><td>macro-planification</td><td colspan="2">micro-planification</td><td>formulation</td></tr>
</table>

Tableau 1. Terminologie liée aux processus rédactionnels

2.2. « Révision »

Dans le modèle de Hayes et Flower (1980), il est indiqué que, durant la phase de « révision », le scripteur relit ce qu'il a écrit afin d'évaluer si son discours correspond à ce qu'il souhaitait exprimer, et d'apporter les corrections qui lui semblent nécessaires. Ce processus est caractéristique de la production écrite, puisque le message n'est pas transmis instantanément au destinataire, cette différence de temporalité entre la production et la réception permettant de revenir sur le texte produit et de le modifier[21] (Lebrave & Grésillon, 2009). Une des caractéristiques de la production écrite, par rapport à l'oral, est que le temps de la réception est autonome du temps de la production, ce qui permet au lecteur de revenir en arrière

Ce concept a fait l'objet de nombreuses études et modélisations, mais sa définition diffère d'un auteur à l'autre (Heurley, 2006; Lindgren & Sullivan, 2006b). Parfois, le terme de « révision » n'est pas employé, comme c'est le cas avec Kellogg (1996), qui évoque la lecture et la correction comme composantes du sous-processus de « contrôle ». Cette ambiguïté pose des problèmes d'un point de vue de la recherche en didactique ainsi que de la recherche fondamentale sur les processus rédactionnels. Ainsi, il a été reproché à Hayes et Flower (1980) de promouvoir le modèle de rhétorique classique, considérant la révision comme l'étape finale de la rédaction, consistant en une relecture-évaluation de l'ensemble de son texte, alors que les processus décrits par ces auteurs sont récursifs et peuvent interrompre les autres processus à n'importe quel moment. Or, Hayes (1991) a rappelé qu'il fallait définir correctement les tâches si l'on souhaitait voir l'évolution de certains processus rédactionnels : ainsi, la « révision » peut ne pas répondre aux mêmes objectifs suivant l'âge ; une amélioration de ce processus pourrait alors ne pas provenir seulement d'une augmentation des capacités des scripteurs, mais aussi d'un changement dans leur définition de cette tâche. Wallace et Hayes (1991) avaient en effet montré que les révisions effectuées par des collégiens étaient de meilleure qualité s'ils avaient reçu des instructions précises sur ce en quoi consistait une « révision ».

Dans son modèle de 1995, Hayes ne mentionne plus la révision comme un processus de base. Il évoque le processus de « traitement du langage », qui à l'écrit correspond à la lecture, et souligne qu'il existe une différence entre la lecture en vue de comprendre un texte, et la lecture en vue de l'évaluer. Par exemple, une activité de compréhension va impliquer de mobiliser ses connaissances pour faire des références, identifier l'idée principale, alors que pour une évaluation il faudra vérifier systématiquement qu'il n'y ait aucune ambiguïté de référence, aucune erreur discursive ou incohérence. En effet, le scripteur est invité à poser des choix à tous les niveaux de la production de texte (constructions

[21] Cette caractéristique n'est cependant pas aussi marquée suivant les situations d'écriture : ainsi, lors d'une discussion par messagerie instantanée le message est reçu presque simultanément à sa production ; il est toutefois toujours possible pour le scripteur de relire son énoncé avant de l'envoyer.

grammaticales, structure du texte, transitions, etc.) que le « réviseur » doit évaluer. Au-delà de l'identification de problèmes de surface, l'évaluateur doit donc chercher à améliorer son texte sur ces différents points. Une fois les problèmes détectés grâce à la lecture, le scripteur s'engage de nouveau dans une activité de réflexion pour résoudre le problème, puis produit la forme correcte pour y remédier. Dès son modèle de révision de 1996, Hayes identifie les trois niveaux visibles dans celui de 2012 : la structure de contrôle, correspondant dans ce cas au schéma de la tâche de révision, les processus fondamentaux que nous venons d'évoquer, et enfin les ressources, constituées de la MDT et de la MLT. Ces composantes diffèrent légèrement dans le modèle de 2012, puisque l'on trouve « l'évaluateur » comme composante des processus cognitifs, qui peut être sollicité avant ou après l'exécution ; la lecture est, elle, intégrée comme ressource, avec la MDT et la MLT.

Le problème de la définition de la révision provient essentiellement des traces laissées par ce processus. En effet, on peut distinguer trois possibilités d'interaction entre évaluation du texte et production :

A. Avant d'écrire la suite de la phrase, le scripteur relit son texte et change d'idée ou de but, modifiant ainsi ce qui avait été prévu initialement.
B. Après avoir écrit une partie du texte, le scripteur relit son texte, et apporte une correction pour remédier à un problème détecté.
C. Après avoir écrit une partie du texte, le scripteur relit son texte en vue de l'améliorer, ne détecte aucun problème et poursuit sa rédaction.

Certains chercheurs limitent la définition de la révision au cas « B », avec la présence nécessaire de « modifications », « changements », « transformations »[22] (e.g. Faigley & Witte, 1981). Les chercheurs qui suivent la définition initiale de Hayes et Flower (cas « B » et cas « C ») mentionnent l'aspect facultatif de la modification du texte, et mettent en avant que l'objectif de la relecture est l'amélioration de la production (Chesnet & Alamargot, 2005). Le cas « A » correspond à ce que Lindgren et Sullivan (2006a) appellent « révisions internes », les révisions « externes » correspondant au cas « B ». Kellogg (1996) évoque également la possibilité d'un « pré-contrôle » avant la production d'une partie du texte.

Le processus de révision, tel que défini par Hayes et Flower en 1980, a fait l'objet de plusieurs modélisations (pour une revue de certains modèles, voir Heurley, 2006). Les principales étapes sont explicitées par Scardamalia et Bereiter (1983) dans le modèle CDO (compare, diagnose, operate) : après avoir comparé le texte écrit avec le texte attendu (représenté mentalement), le scripteur identifie l'origine des différences : s'il les trouve, il peut soit choisir de changer le plan, apportant des modifications au texte attendu, soit choisir la stratégie appropriée pour modifier le texte réel. Si le scripteur est en échec à l'une des étapes, il retourne au processus qui avait été interrompu. En 1987, Hayes, Flower, Schriver,

[22] En anglais, le terme editing est fréquemment utilisé

Stratman et Carey modélisent les différentes étapes du processus de révision, et les mettent en relation avec la composante « connaissances » : l'évaluation est possible si le texte est mis en relation avec les buts et contraintes fixés ; elle donne lieu à un nouveau type de connaissance, i.e. la représentation du problème ; enfin, l'étape de modification est liée au répertoire des moyens/fins et aux procédures d'améliorations connues par le scripteur, ces connaissances pouvant être enrichies par la modification effectuée. Dans ce modèle, une distinction est faite entre « réécrire » (rewriting) et « réviser » (revising) : d'après les auteurs, il s'agit des extrêmes d'un continuum entre la réécriture complète du texte, et la modification de quelques éléments dans le but de conserver au maximum ce qui a été déjà écrit, ce qui n'est possible que si le problème a été bien identifié et que le scripteur possède les moyens nécessaires pour y remédier. Alors que ce modèle ne traitait pas du lien entre la révision et les autres processus, Butterfield, Hacker et Albertson (1996) ont modélisé le processus de révision en lien avec l'environnement et les systèmes cognitifs et métacognitifs engagés de la production. Ainsi, les différentes étapes de la révision (détection de problème, sélection des stratégies, etc.) se déroulent en MDT, en étant en lien avec la MLT au sein de laquelle le scripteur conserve ses connaissances, ses stratégies et la représentation du texte en train d'être révisé. La MLT contient également une composante métacognitive, qui permet de coordonner les connaissances et stratégies afin d'assurer la révision. Enfin, dans l'environnement, on retrouve les considérations rhétoriques (e.g. thème, destinataire) et le texte qui est en train d'être révisé.

Plusieurs types d'informations peuvent être relevés quant aux révisions apportées dans le texte. Tout d'abord, il est possible d'identifier l'objet de la révision. Dans leur taxinomie, Faigley et Witte (1981) distinguent ainsi les changements de surface des changements portant sur le texte. Lindgren et Sullivan (2006a) conservent une typologie similaire en distinguant les révisions formelles (erreur de frappe, orthographe, grammaire, ponctuation, forme du texte, changement lexical avec préservation du sens) des révisions conceptuelles (microstructure, macrostructure, adaptation au destinataire, équilibre, adéquation au thème). Même si la distinction entre révision « formelle » et « conceptuelle » n'est pas toujours explicitée, les chercheurs signalent généralement ainsi l'objet de la révision (orthographe, sémantique, organisation). Il est cependant difficile d'affirmer que les corrections orthographiques et lexicales ne sont toujours que le reflet d'une révision « de surface » (David, 1994) ; il peut y avoir ainsi un décalage entre la marque laissée et le processus cognitif sous-jacent. Lindgren et Sullivan (2006a) signalent également les révisions situées au « point d'inscription », c'est-à-dire lorsque le scripteur modifie ce qu'il est en train d'écrire ou ce qu'il vient juste d'écrire. Dans ce cas, comme il n'est pas possible d'accéder à la représentation initiale du texte, il est difficile de savoir s'il s'agit d'une modification mineure, portant par exemple juste sur un choix lexical, ou si le scripteur a eu une nouvelle idée. Chanquoy (2001) suggère par ailleurs de préciser si la correction est correcte (corrige une erreur dans le texte), erronée (créé une erreur) ou neutre (ne corrige pas d'erreur, n'en crée pas).

Il est en outre possible d'envisager les révisions du point de vue des opérations effectuées, en indiquant par exemple s'il s'agit d'un ajout, d'une suppression, d'un remplacement ou d'un déplacement (Doquet, 2003). Si cette typologie peut paraitre intuitive, son application soulève cependant des questions. Par exemple, l'ajout est identifié dans les manuscrits lorsqu'un élément est inscrit en dehors du texte principal (entre deux lignes, dans la marge) ; ce critère typographique n'est pas valable dans le cas d'un texte écrit par ordinateur : il sera alors nécessaire de recourir à la temporalité de l'écriture pour identifier les segments insérés au sein du texte déjà écrit. Cependant, on peut également dire que lorsqu'on écrit une phrase, on l'« ajoute » au texte déjà écrit, même si c'est à la suite de ce texte ; Doquet (2003) considère que dans ce cas on peut également parler d'ajout. La notion de « remplacement », qui consiste en une suppression et un ajout, pose également des problèmes de définition : il est alors possible de l'identifier dès lors qu'il y a une équivalence syntagmatique entre le terme supprimé et celui ajouté, et que les deux opérations se succèdent au niveau chronologique. De manière générale, on peut considérer que toute révision équivaut à substituer un segment à un autre (Fabre, 1991), même si le segment initial (cas de l'ajout) ou final (cas de la suppression) correspond à la variable « Ø » (Lebrave & Grésillon, 2009)

Ainsi, de nombreux éléments peuvent être pris en compte lorsque l'on s'intéresse à la révision : type d'opération, chronologie des opérations, type d'unité corrigée, contenu de la correction (Rey-Debove, 1981). Dans le cadre de notre recherche, nous utiliserons le terme de « révision » pour parler des opérations ayant laissé une trace de retour sur le texte déjà écrit.

3. La gestion des processus

La section précédente nous a permis d'identifier les processus rédactionnels principaux, tout en soulignant la diversité des définitions et terminologies adoptées. Notre travail portant sur la gestion temporelle de ces processus, nous allons présenter quelques modèles liés à la mémoire de travail (MDT), qui sont un cadre nécessaire à la compréhension de la gestion des processus cognitifs. Nous indiquerons en outre de quelles manières les interactions entre les processus rédactionnels peuvent être envisagées.

3.1. Les contraintes de capacité cognitive

Les activités complexes permettant la production écrite doivent être intégrées par le scripteur en temps réel, une des difficultés de cette gestion étant que sa capacité de traitement est limitée (Bourdin, 2002; Fayol, 2005). La gestion des processus nécessite de prendre en compte les différentes composantes de la mémoire, qui d'après Piolat (2004) renvoie dans le langage quotidien à une « fonction de stockage mental de l'information », ainsi que la manière dont les ressources attentionnelles sont gérées.

Les études menées en psychologie cognitive ont permis de mieux comprendre certains fonctionnements de la mémoire (pour une revue, voir Chanquoy & Alamargot, 2002; Piolat, 2004). Il faut tout d'abord distinguer deux grands types de mémoire : la mémoire à long terme (MLT) et la mémoire à court terme (MCT) ; lorsque l'on parle de la mémoire dans le cadre d'une activité, la notion de mémoire de travail (MDT) est généralement utilisée. La MLT contient entre autres un ensemble de connaissances dans lequel le scripteur va puiser aux différentes étapes de sa rédaction. Il a également été proposé que certaines informations de la MLT restent davantage accessibles au cours de la rédaction, leur récupération étant alors plus facile : c'est ainsi qu'a été introduit le concept de la mémoire de travail à long terme (Ericsson & Kintsch, 1995; McCutchen, 2011). La MDT a été particulièrement décrite par Baddeley et Hitch, dès 1974. Selon eux, celle-ci est constituée du calepin visuo-spatial, qui traite les informations visuelles, et de la boucle phonologique, qui s'occupe des informations auditives, ces deux composantes étant gérées par l'administrateur central. À cela a été ajouté en 2000 le buffer épisodique, dont le rôle est notamment de faire le lien entre la MLT et la MCT. Kellogg (1996, 1999) a cherché à déterminer quelles composantes de la MDT étaient utilisées pendant les différents processus de rédaction. Hormis pour l'exécution motrice chez les experts, qu'il considère comme entièrement automatique, les ressources de l'administrateur central sont indispensables à chaque étape pour gérer l'allocation des ressources. Certains processus font ensuite plutôt appel à l'une ou à l'autre des composantes : alors que la planification, de par sa nature conceptuelle et pré-linguistique, mobilise le calepin visuo-spatial, la formulation et la relecture se servent plutôt de la boucle phonologique. Cependant, une étude récente montre que les recherches sur cette question se poursuivent (Kellogg et al., 2016).

La MDT, qui est utilisée par les scripteurs pour stocker des informations et pour effectuer des traitements mobilisant des processus cognitifs, apparait cependant comme une ressource limitée. Just et Carpenter (1992) se sont servis de ce modèle de la MDT pour expliquer le développement de la compréhension et les différences interindividuelles observées. Ils ont développé la « théorie capacitaire », qui postule une limitation des ressources cognitives, conduisant celles-ci à être allouées de manière inégale à tous les processus. McCutchen (1996) a adapté cette théorie à la production écrite. Ainsi, les ressources de la mémoire étant limitées, un cout cognitif élevé pour certains traitements pourra diminuer les ressources disponibles pour d'autres : par exemple, si l'exécution motrice demande un très grand effort, la planification ou la révision pourront se trouver affectées. Les processus peuvent fonctionner en parallèle tant que les ressources cognitives ne sont pas épuisées. Le ralentissement de la vitesse d'écriture ou l'allongement des pauses seraient en revanche dus à une surcharge cognitive. Cette surcharge cognitive est due à la non-automatisation de certains processus : dans le cadre d'une production ne demandant aucune hésitation quant à la planification, à la formulation, à l'orthographe, et pour laquelle l'exécution est complètement automatisée, l'on pourrait imaginer une production totalement fluide, sans pauses ni variation de la vitesse d'écriture. D'ailleurs, Hayes et Chenoweth (2006) ont montré que, lorsque des scripteurs devaient simplement copier un texte d'une fenêtre d'ordinateur à une autre il n'y avait pas de bursts, c'est-à-dire de périodes d'exécution motrice séparées par des pauses ; la variation du rythme d'écriture n'est donc pas due au processus de transcription lui-même, du moins dans les cas où celui-ci est automatisé. Généralement, lorsqu'il y a une surcharge cognitive à cause de l'exécution et de la formulation, c'est ce dernier processus qui sera privilégié : le scripteur va alors ralentir ou s'arrêter (Brown, McDonald, Brown & Carr, 1988). Une autre stratégie consisterait à continuer à produire du texte même si le contenu n'est pas satisfaisant, puis à réviser en profondeur ce qui vient d'être écrit ; cette stratégie est cependant plus rarement relevée, bien qu'il soit possible de lier la longueur des périodes d'exécutions avec la quantité de révisions effectuées (Baaijen, Galbraith & de Glopper, 2012). Le partage des ressources cognitives est ainsi visible à travers l'interaction entre les différents niveaux de rédaction : pour les personnes peu habituées à écrire, les processus de bas-niveau demandent beaucoup d'efforts cognitifs, affectant le nombre et la nature des autres processus rédactionnels engagés.

Cette théorie capacitaire apparait donc comme un cadre pertinent pour appréhender la temporalité de l'écriture : le cout de chaque processus et la gestion en parallèle ou en succession de ces processus sera notamment reflétée dans les variations chronométriques relevées au long de la rédaction.

3.2. Interactions entre les processus rédactionnels

Plusieurs types de relations peuvent être envisagés entre les processus. Tout d'abord, si l'on considère le passage d'un niveau à l'autre, il peut s'agir soit de traitements séquentiels, c'est-à-dire que l'output d'un niveau constitue l'input du niveau suivant, l'information circulant dans « un sens », soit

d'un traitement connexionniste, le passage d'un niveau à l'autre étant bi-directionnel. En outre, les traitements peuvent être effectués de manière plus ou moins parallèle, donnant parfois lieu à des interactions entre les différents niveaux de traitement.

Plusieurs modèles ont tenté de décrire de manière précise les relations entre les différentes étapes de la « formulation », c'est-à-dire la sélection des constructions syntaxiques, des lexèmes, et des formes phonologiques. Nous n'entrerons pas dans le détail de ces modèles, qui se situent à un niveau inférieur (syntagme, mots, graphèmes) à ceux étudiés dans le cadre de notre recherche. Ils ont généralement été formalisés à partir des données chronométriques et de l'analyse des erreurs de production. Les modèles sériels classiques (Bock & Levelt, 1994; Levelt, 1989, 1999; Roelofs, 1997) stipulent que la sélection d'un lexème ne peut se faire qu'à l'issue de la sélection d'un lemme. Des interactions entre les différents niveaux peuvent cependant parfois se produire s'il l'on considère qu'il s'agit d'un modèle en cascade, dans lequel un niveau (par exemple la sélection du lexème) peut être activé même si toute l'information n'a pas fini d'être totalement traitée au niveau précédent (Caramazza, 1997). Dans ce cas, plusieurs niveaux sont activés en parallèle, pouvant provoquer des interactions entre eux. Les interactions sont encore plus importantes dans les modèles connexionnistes (Dell, 1986), stipulant que les différents niveaux traitant un même item sont activés en même temps : ainsi, la coactivation du lemme et de la forme du mot peut expliquer que les caractéristiques phonologiques d'un mot influencent le choix lexical.

À l'échelle de la production écrite d'un texte, la vision extrême d'un modèle sériel stipulerait qu'à la planification succèdent la formulation puis la révision. Ce fut d'ailleurs le reproche souvent adressé au modèle de Hayes et Flower (1980) : certains détracteurs assimilaient ce modèle à celui de la rhétorique classique en pensant que les processus présentés se succédaient les uns aux autres. En effet, les modélisations issues de la rhétorique classique se donnaient pour but de fournir un « modèle », c'est-à-dire un exemple, et pour cela présentaient une suite d'étapes dont l'ordre était immuable (Plane, 2004). En réalité, Hayes et Flower ont tout de suite souligné la récursivité des processus rédactionnels : chacun de ces processus peut en interrompre un autre à tout moment de la production, d'où la nécessité d'avoir une composante de « contrôle » (monitoring) pour assurer la passage d'un processus à l'autre. Les modèles plus récents (Chenoweth & Hayes, 2001; Hayes, 2012) soulignent quant à eux la fonction de l'évaluateur, qui contrôle fréquemment ce qui ressort des autres modules (proposeur, formulateur, transcripteur) et permet que l'information soit re-traitée dans l'un des modules si nécessaire.

La théorie capacitaire suggère que certains processus peuvent être traités en parallèle, tant qu'il n'y a pas de surcharge cognitive, ce qui permet notamment de rendre compte d'une certaine fluidité dans la production (Olive, 2002). Plus particulièrement, les processus de planification, formulation et révision peuvent être effectués pendant l'exécution motrice, lorsque celle-ci est suffisamment automatisée, ce qui permet de maintenir une certaine fluence d'écriture (Olive, 2014). En effet, l'étude d'Olive, Alves et Castro (2009) montre que, pendant la transcription graphique, d'autres processus rédactionnels sont

engagés : en même temps qu'ils écrivaient, il arrivait aux scripteurs de réfléchir à la formulation ou à la planification ; par moments, ils effectuaient même des révisions en même temps que la transcription. En ayant recours à la double tâche, Olive et Kellogg ont montré dans leur étude de 2002 que les temps de réactions des adultes rédigeant avec leur écriture familière étaient plus élevés lors de l'exécution motrice au cours de la tâche de rédaction que lors de la tâche de copie : cela suggérait que des processus supplémentaires, tels que la planification ou la formulation, se déroulaient en parallèle de l'exécution. En revanche, les enfants et les adultes en situation d'écriture non familière (écriture en majuscules) avaient un temps de réaction aussi rapide lors de la composition de texte que lors de la copie, les processus de haut-niveau étant sans doute mobilisés principalement durant les pauses, puisque l'exécution motrice n'était pas parfaitement automatisée. Le traitement parallèle des processus peut aussi s'étendre à la lecture du texte déjà écrit : ainsi, Alamargot, Dansac et Chesnet (2007) ont relevé des cas de relecture en même temps que la transcription. Ce type de relecture, qui conduisait à une diminution de la vitesse d'écriture, était plus fréquent pour les scripteurs ayant une plus grande fluence lexicale et une plus grande fluence graphomotrice. Outre le traitement en parallèle des processus, Olive (2014) a proposé un modèle en cascade pour les processus de conceptualisation, formulation et exécution, comme le montre la Figure 6 .

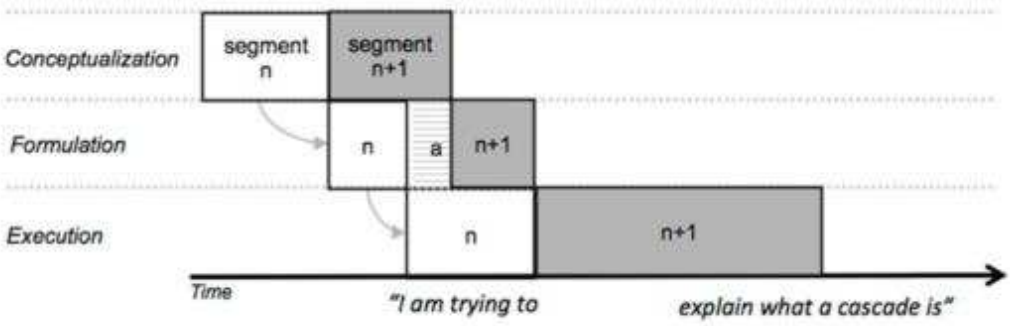


Figure 6. Modèle en cascade de Olive (1994)

Cependant, le fait que la formulation succède à la micro-planification, c'est-à-dire que les idées soient « traduites » en mots, est une conception à la fois répandue et remise en cause. Ainsi, on peut se demander si une planification peut être seulement conceptuelle, ou si cette conceptualisation nécessite une mise en mot. Chenoweth et Hayes (2001) reconnaissent d'ailleurs que le lien entre « proposeur » et « formuleur » n'est pas uni-directionnel : si un concept est plus facile à formuler, il peut être choisi dans le proposeur. Cependant, la distinction entre ces deux niveaux permet d'expliquer certains résultats, comme le changement de stratégie lorsque l'on écrit en langue étrangère : c'est alors bien la formulation qui est rendue difficile, et non la conceptualisation. On pourrait concevoir des échanges entre ces deux niveaux : ainsi, le scripteur peut souhaiter introduire un concept en sachant par exemple quel prédicat et quels noms il veut utiliser ; puis il conceptualise les différentes composantes de sa phrase, formule en introduisant tous les lexèmes nécessaires à une bonne construction syntaxique (e.g. les déterminants, auxiliaires), et peut modifier les lemmes choisis initialement pour davantage de cohésion, par exemple

en pronominalisant un terme ou en choisissant une désignation plus cohérente avec la signification désirée (van Dijk & Kintsch, 1983). On peut également imaginer qu'un lexème soit remplacé par un autre car l'orthographe n'est pas assez bien connue : un scripteur ayant peur de faire des erreurs d'orthographe choisira alors un mot dont la consistance orthographique est plus élevée, c'est-à-dire pour lequel le choix des graphèmes sera plus facile. Des interactions peuvent également avoir lieu au niveau de la macro-planification. Ainsi, Galbraith (1999, 2009) propose que, lors de l'écriture, le scripteur acquiert de nouvelles connaissances : ces nouvelles connaissances peuvent alors modifier le plan initialement prévu, le texte transcrit influençant donc la macro-planification.

Ainsi, les interactions entre les processus témoignent que la production écrite est plus complexe qu'une simple succession linéaire de traitements. Il est donc possible de s'interroger sur l'aspect temporel de la gestion de ces processus cognitifs au cours de la production écrite d'un texte.

3.3. Gestion temporelle

3.3.1. Planification et formulation

Le processus de planification peut se dérouler à divers moments de l'écriture. Il est ainsi possible de faire une première macro-planification avant de commencer à écrire : c'est généralement durant ce que l'on appelle la « pause de pré-écriture ». Cette pause initiale est plus ou moins longue suivant le style du scripteur, mais également suivant des facteurs liés au contexte d'écriture ; par exemple, cette pause de pré-écriture est plus courte pour un texte narratif qu'argumentatif (Beauvais et al., 2011). En outre, la temporalité de l'écriture du texte peut varier en fonction de la longueur de cette pause initiale : par exemple, si le temps de pré-écriture est plus court, le scripteur peut être amené à relire plus fréquemment l'amorce et à faire des pauses plus longues (Epting, Gallena, Hicks, Palmer & Weisberg, 2013).

Si la formulation est le processus activé le plus fréquemment, la planification est également activée tout au long du texte (Kellogg, 1987). Cependant, en divisant le texte en trois parties, Breetvelt, van den Bergh et Rijlaarsdam (1994) ont mis en avant que le processus de génération d'idées était moins fréquent à la fin du texte. Fortier et Préfontaine (1994) ont quant à eux distingué l'organisation conceptuelle (qui équivaudrait à la macro-planification) et l'organisation séquentielle (micro-planification), et ont montré que ces processus étaient plus longs au milieu de la production, tandis que l'organisation conceptuelle était moins longue à la fin de la production, et l'organisation séquentielle moins longue au début de la composition.

Au niveau des rapports entre micro-planification et formulation, Chenoweth et Hayes (2001) suggèrent que le proposeur génère le matériel pré-linguistique, puis le transmet au formulateur qui le stocke dans le buffer articulatoire en vue de l'évaluer, qui le transmet au transcripteur pour que l'énoncé soit ajouté au texte déjà écrit. Cette façon de procéder correspond notamment au profil des scripteurs

qui planifient une phrase, puis la transcrivent et la révisent, avant de passer à la phrase suivante (Flower & Hayes, 1980).

La gestion temporelle de ces processus donne lieu à plusieurs interrogations : quel est l'empan de la planification ? Dans quelle mesure ces processus interagissent-ils ? Comment sont-ils traités par rapport à l'exécution motrice ? Quels facteurs rendent ces processus plus ou moins couteux, donc plus ou moins longs à être traités ?

3.3.2. Révision

Alors que les corrections ont tout d'abord été analysées sur le produit final dans le cadre de la critique génétique, l'accès à des outils d'enregistrement de l'écriture a permis de mettre au premier plan la dimension temporelle de ces révisions : cette temporalité peut être envisagée de manière relative, lorsque les opérations d'écriture sont reconstruites dans leur succession temporelle, mais il peut également s'agir d'une temporalité absolue, si elle est considérée par rapport au début de l'activité d'écriture (Leblay, 2009).

Tout comme les autres processus, la révision (prise au sens large : relecture et modifications) peut intervenir à plusieurs moments de la production, et pas seulement à la fin de la rédaction, pendant un temps où l'ensemble de la production est relu et vérifié. Au niveau de l'écriture du texte dans son ensemble, la révision est le processus le moins souvent activé (Kellogg, 1987), et il l'est davantage à la fin qu'en début d'écriture (Breetvelt et al., 1994). La fréquence des révisions peut en outre dépendre du medium d'écriture, les élèves corrigeant davantage leur texte lorsqu'ils composent par ordinateur qu'à la main (Doquet, 2006). Concernant la temporalité des révisions, Severinson, Eklundh et Kollberg (2003) soulignent qu'il existe des cas où les révisions se succèdent. Cela peut arriver lors d'une révision en fin de texte où différentes corrections sont apportées, mais également lorsqu'un changement effectué implique de nouvelles modifications, donnant lieu à des « sessions de révisions ».

Par ailleurs, Hayes et Flower (1980) distinguent notamment le processus d'editing, qui est automatique et consiste en de brefs épisodes de retour sur le texte interrompant les autres processus, et le reviewing, lorsque le scripteur décide de consacrer un temps pour examiner son texte en vue de l'améliorer. Ainsi, Kaufer, Hayes et Flower (1986) ont relevé que les scripteurs interrompaient fréquemment leur production afin de corriger une erreur, par exemple grammaticale. Fayol (1997) signale également que des retours sur le texte, même sans modification, sont très fréquents et systématiques même chez les scripteurs les plus jeunes ; ces relectures portent généralement sur le mot qui vient d'être écrit, ou du moins sur la proposition ou phrase en train d'être écrite. Une des difficultés de l'étude de la révision, liée à la complexité des définitions données, vient du fait que celle-ci ne laisse parfois pas de trace : ainsi, Kellogg (1996) indique que la vérification peut certes avoir lieu après l'écriture de la phrase, mais également avant le début de la transcription. C'est en effet le rôle de l'évaluateur qui juge dans quelle mesure le message produit par le formulateur est correct et peut être

mis par écrit. En outre, il n'est pas possible d'associer systématique « relecture » et « révision » : Fortier et Préfontaine (1994) ont par exemple mis en évidence la présence de relecture du plan ; dans ce cas, la relecture est associée à la planification, et non à la révision.

La question de la gestion temporelle du processus de révision se pose également à l'échelle de la phrase. Dans un cadre expérimental plus contraint et limité à la production d'une phrase, Quinlan, Loncke, Leijten et Van Waes (2012) ont montré que les scripteurs préféraient finir leur phrase avant d'effectuer une correction, d'autant plus si de nombreuses informations devaient être gardées en mémoire. Cela témoigne du besoin de se libérer d'une certaine charge cognitive avant d'entreprendre un processus de révision. En effet, lors de l'écriture d'une phrase isolée, les participants mobilisent notamment la boucle phonologique afin de garder en mémoire les mots servant à compléter la phrase ; dans le cadre de la production d'un texte, les éléments à garder en mémoire sont encore plus nombreux : outre les mots, il faut mémoriser la structure de la clause en train d'être écrite, les liens référentiels avec les phrases précédentes, etc. La mobilisation de la boucle phonologique ainsi que du calepin visuo-spatial lors de la mémorisation peut rendre difficile une tâche de révision en parallèle, cette dernière devant puiser dans les mêmes ressources cognitives. On pourrait alors imaginer que le fait d'observer où les révisions sont initiées peut nous renseigner sur les unités de production, puisque le scripteur attendrait d'avoir transcrit la séquence planifiée avant d'apporter une modification.

4. Différences interindividuelles

Lors de la présentation de la gestion des processus, il est arrivé à plusieurs reprises d'évoquer des différences possibles suivant les individus. Comme nous allons le montrer, ces différences peuvent être dues au degré de développement de l'expertise, notamment par rapport à l'âge des scripteurs, mais également à d'autres facteurs de variation entre individus, ou à des styles d'écriture dont les scripteurs ont plus ou moins conscience.

4.1. Développement de l'expertise

Le modèle initial de Hayes et Flower (1980) ne permettait pas de rendre compte de la production écrite d'un point de vue développemental. En revanche, des études menées ultérieurement ont permis de compléter ce modèle, et de souligner l'importance de la prise en compte de certaines composantes. Certes, certains éléments de la production du langage sont communs à l'écrit et à l'oral, comme le fait de générer des idées à partir de la MLT et de les formuler à travers des choix syntaxiques et lexicaux ; mais l'enfant doit tout de même acquérir les spécificités liées à la modalité écrite, ce qui concerne le contenu linguistique, comme nous l'avons vu précédemment, mais aussi les processus. Parmi les différences entre les modalités, Fayol (1997) souligne la dimension proprement graphique du texte, mais également l'absence de pression communicative ; il met cependant en garde sur l'aspect schématique de la dichotomie oral / écrit, proposant par exemple de prendre également en compte l'aspect formel / non formel de la communication. À partir de ces considérations, Chanquoy et Alamargot (2003) font ressortir quatre caractéristiques du langage écrit : les dimensions graphique (transmission du message par l'intermédiaire de lettres), orthographique (respect des conventions du système orthographique de la langue), textuelle (utilisation du texte déjà écrit pour produire un texte cohérent) et communicationnelle (hypothèses sur le lecteur). Or, l'expertise concernant ces différentes dimensions se développe.

Tout d'abord, parvenir à une transcription graphique automatique requiert à la fois la présence et l'accès à la représentation précise des lettres, mais également l'accès à une planification motrice permettant de produire ces lettres. Or, les enfants reçoivent assez peu d'enseignement explicite pour améliorer ces compétences, ce qui fait que, pour beaucoup, le geste graphique reste un problème, réduisant notamment la qualité de leurs compositions (Berninger et al., 1997). Ainsi, que ce soit dans une tâche de rappel de nombre ou d'histoire, les adultes ont les mêmes performances à l'écrit qu'à l'oral, alors que les enfants obtiennent de meilleurs résultats à l'oral (Bourdin & Fayol, 1994; Bourdin, Fayol & Darciaux, 1996). Hayes et Berninger (2010) ont également montré que, chez les enfants, moins d'idées sont exprimées à l'écrit par rapport à l'oral, et en tapant leur texte sur un ordinateur plutôt qu'en l'écrivant à la main. L'impact des compétences de transcription (formation des lettres et orthographe) sur la capacité de combiner des phrases entre elles, ou le fait que, dans le cas de la production d'un texte, la complexité des phrases soit affectée par le mode de transcription (Berninger, Nagy & Beers, 2011),

montrent qu'une forte mobilisation des ressources cognitives pour le processus de transcription se fait alors au détriment de l'encodage syntaxique.

Une étude de cas analysant les mouvements oculaires sur un corpus développemental a mis en évidence une évolution quant au schéma de lecture globale et locale pendant la composition d'un texte (Alamargot, Plane, Lambert & Chesnet, 2010), avec une relecture plus locale pour les plus jeunes. Cette évolution de l'utilisation du texte déjà écrit peut être mise en lien avec la dimension communicationnelle. En effet, les auteurs lient ces résultats aux différentes stratégies de planification identifiées par Bereiter et Scaradamalia (1987) et Kellog (2008). Bereiter et Scardamalia (1987) ont effectué une distinction entre deux types d'utilisation des connaissances présentes dans la MLT. Le premier, connu sous le nom de « connaissances rapportées » (knowledge telling), consiste en une extraction des connaissances selon un ordre relativement chronologique. Il s'oppose à la stratégie de « connaissances transformées » (knowledge transforming), qui nécessite une réorganisation des connaissances pour obéir à un ordre logique. Cette distinction peut par exemple être pertinente dans le cas d'une distinction entre discours expositif, relevant plutôt de la stratégie des connaissances transformées, et un discours narratif qui utiliserait la stratégie des connaissances rapportées. D'un point de vue développemental, on peut considérer que les enfants utilisent dans un premier temps davantage la stratégie de connaissances rapportées, moins couteuse au niveau cognitif. Dans ce cas, le scripteur va plutôt prendre en compte la cohérence locale et donc effectuer moins de relecture globale. Kellogg (2008) identifie une troisième stratégie, celle des « connaissances retravaillées » (knowledge-crafting), caractéristique des scripteurs experts. Les stratégies de connaissances transformées et retravaillées ne diffèrent pas par l'accès aux connaissances ni par les interactions entre les trois processus rédactionnels, mais par le fait que, lors des révisions, le scripteur expert prend non seulement en compte ses propres représentations mentales, mais également celles de son interlocuteur.

La gestion temporelle des processus rédactionnels a été évoquée par Berninger et Swanson (1994). Chez les scripteurs les plus jeunes, la planification se fait principalement au fur et à mesure de l'écriture, et les révisions relevées se font généralement au cours de l'écriture ; la génération de texte lors de la formulation se fait tout d'abord au niveau du mot, puis de la phrase, puis du texte. Par la suite, l'exécution motrice est davantage automatisée, ce qui permet aux enfants de stade « intermédiaire » d'utiliser des structures linguistiques plus appropriées au type de discours ; c'est la période où émergent le processus de planification en début de texte, ainsi que la révision après l'écriture du texte. Pour les scripteurs les plus avancés, les différences de gestion se font notamment en fonction de l'empan de la MDT, dont les performances connaissent une augmentation linéaire de 4 à 15 ans (Gathercole, Pickering, Ambridge & Wearing, 2004), et en fonction des connaissances métacognitives.

Même si la majorité des études « développementales » concernent les enfants et adolescents, ce n'est pas pour autant que le développement de l'expertise est achevé au lycée. Il est en effet possible

d'étudier les productions d'étudiants en master ou en doctorat, ou des rédacteurs professionnels de textes procéduraux ou des écrivains (Alamargot & Lebrave, 2010; Alamargot et al., 2010).

4.2. Profils de scripteurs

Des différences interindividuelles peuvent apparaitre, même entre adultes. Tout d'abord, Berninger, Fuller et Whitaker (1996) soulignent l'importance de prendre en compte l'empan de la MDT, liée à la génération de texte, ainsi que de la MCT, liée à la transcription. Par exemple, Alamargot, Caporossi, Chesnet et Ros (2011), grâce à l'analyse du texte produit, des mesures chronométriques et des mouvements oculaires, ont montré que les participants dont la MDT avait la capacité la plus élevée faisaient des pauses plus longues, et différaient dans leur stratégie de lecture du texte source, en mettant à profit leurs ressources cognitives pour mieux prendre en compte leur destinataire.

La fluence d'écriture et la qualité des textes sont également en partie expliquées par les compétences en transcription, par exemple l'agilité des mouvements des doigts, ou les compétences en orthographe (Berninger & Swanson, 1994). L'influence de l'exécution motrice sur les processus de rédaction pourrait sembler assez faible chez les adultes, le geste graphique étant relativement automatisé chez ces derniers. Cependant, Connelly, Dockrell et Barnett (2005) ont montré une corrélation entre la fluence d'écriture d'étudiants et la qualité de leur travail lors d'un examen, si celui-ci était effectué sous pression temporelle ; la qualité de la prise de note des étudiants est également liée à leur fluence d'écriture (Peverly & Sumowski, 2012). Ainsi, même avec des scripteurs experts, la variable de l'aisance lors de l'exécution motrice ne doit pas être négligée. De même, le cout de gestion de la production écrite provoque une diminution du nombre d'idées élaborées par des adultes à partir de mots non reliés entre eux, ce qui n'est pas le cas dans le cas d'une production orale (Bourdin & Fayol, 2002). Lorsqu'il n'y a pas de pression temporelle, il est possible que la qualité du texte ne soit pas affectée, mais dans ce cas la gestion temporelle globale de l'écriture est modifiée : ainsi, les scripteurs les plus lents à l'ordinateur produisent moins de mots pendant les périodes d'exécution et passent plus de temps en pauses, celles-ci étant plus nombreuses, et plus longues dans tous les contextes hormis en début de phrase et après la virgule (Alves, de Sousa & Strömqvist, 2007). À l'inverse, si la stratégie de gestion n'est pas modifiée, la qualité du texte peut s'en trouver affectée (Alves, Castro & Olive, 2008).

Des stratégies de gestion des processus, plus ou moins conscientes, ont pu être mises à jour : par exemple des scripteurs planifiant longtemps avant de commencer la mise en texte, d'autres planifiant à chaque phrase, les périodes de révision pouvant aussi différer suivant les personnes. Baaijen et al. (2012) ont par exemple montré que lorsqu'il y avait des phrases précédées de pauses longues, les périodes d'exécution étaient plus courtes, et il y avait moins de révisions. L'étude des périodes de planification et de révisions ont notamment donné lieu à des classifications. En se référant aux célèbres compositeurs, Boehm (1993) a par exemple distingué les Mozartians, qui planifient au début, puis récupère au fur et à

mesure ce qui a été planifié et révisent le texte au long de l'écriture, et les Beethovians, qui génèrent les idées au fur et à mesure de l'écriture et révisent à la fin de leur composition ; cela correspond à ce que Galbraith (1992) a nommé respectivement méthodes « classiques » et « romantiques ». Cette distinction rejoint dans une certaine mesure celle mise à jour par Hay (cité dans Alamargot & Lebrave, 2010) dans le domaine de la génétique textuelle, entre l'« écriture à programme », caractérisée par une planification importante avant le début de l'écriture, et l'« écriture à processus » où le texte est construit par l'auteur au moment où il est en train d'écrire. Là encore, des exemples d'auteurs célèbres peuvent être donnés, grâce à l'étude de leur manuscrit : Zola est un exemple d'auteur d'« écriture à programme », l'« écriture à processus » étant privilégiée par exemple par Proust. L'établissement de profil n'est pas réservé aux écrivains professionnels : chez les enfants en école primaire, il est possible de distinguer le « mode de l'accumulation », caractérisé par une longue pause en début d'écriture et des pauses régulières à la fin des énoncés longs, et le « mode de la récurrence », où le scripteur revient fréquemment sur son texte après avoir écrit un énoncé, puis continue en produisant l'énoncé suivant, qu'il relit et modifie si nécessaire (Doquet, 2006). Ces distinctions sont les extrêmes d'un continuum, et l'on voit que la dichotomie n'est pas exactement la même suivant les études. La classification peut en outre dépendre de la situation d'écriture : par exemple, en demandant à leurs participants de produire deux versions intermédiaires avant la version finale, Van Waes et Schellens (2003) ont distingué cinq profils de scripteurs à partir des temps de planification et de la fréquence des révisions en fonction des étapes d'écriture. De plus, ces auteurs n'hésitent pas à reconnaitre que certains scripteurs ne peuvent pas être assimilés à un profil particulier, et créent le profil des « scripteurs dans la moyenne » (average writers), qui puisent des caractéristiques de plusieurs profils. On voit en outre que ces profils sont également affinés en fonction du matériel d'étude. En effet, les généticiens se basent essentiellement sur la présence de versions successives d'un même texte et sur les corrections visibles ; les études enregistrant la temporalité de l'écriture s'appuient également sur la longueur des pauses et le lieu d'origine des révisions ; enfin, il est actuellement possible d'enrichir ces profils par le comportement lié à la lecture, grâce à l'enregistrement des mouvements oculaires (Dragsted & Carl, 2013).

Des gestions temporelles différentes peuvent donc exister entre les scripteurs, même au sein d'un groupe qui pourrait apparaitre homogène. Ainsi, dans une étude de cas, Fortier et Préfontaine (1994) ont relevé les pauses supérieures à 1 seconde de quatre participants : un des participants faisait peu de pauses mais beaucoup de pauses très longues (plus de 20 secondes), alors que la moyenne indiquait que 60% des pauses étaient inférieures à 5 secondes. La différence de gestion temporelle était aussi visible dans la localisation des pauses longues au sein du texte. Cela est donc à prendre en compte lors des études, l'utilisation des moyennes pouvant masquer de fortes différences.

5. Conclusion

Les modèles présentés permettent de mieux comprendre l'ampleur des travaux concernant la production écrite, ainsi que la difficulté de leur réalisation.

Concernant les processus rédactionnels, des termes issus du modèle initial sont employés couramment par de nombreux chercheurs : planification, formulation, révision. Nous distinguerons pour notre part la macro-planification (choix conceptuels concernant la structure du texte et ses principales composantes), la micro-planification (choix sémantiques à un niveau local) et la formulation (choix syntaxiques et lexicaux permettant d'écrire le contenu de la micro-planification sous la forme d'un énoncé grammaticalement correct). Les différentes définitions liées à la révision concernent principalement la nécessité ou non de la présence de modifications, ainsi que l'aspect plus ou moins systématique de ces retours sur le texte. Dans notre étude, nous utiliserons le terme de « révision » pour désigner les opérations ayant laissé une trace de retour sur le texte déjà écrit.

La théorie capacitaire nous parait être un cadre pertinent pour appréhender la temporalité de l'écriture. En effet, il sera considéré que les variations chronométriques relevées au long de la rédaction reflètent le cout de chaque processus et la gestion en parallèle ou en succession de ces processus. Même si l'aspect sériel de certains traitements peut permettre de poser des hypothèses quant à la gestion temporelle des processus, nous tiendrons compte de la complexité des interactions entre les processus rédactionnels, certains processus de « bas-niveau » (phonologiques, lexicaux, syntaxiques) n'étant pas systématiquement traités après les processus de « haut-niveau » (sémantiques).

Ces cadres conceptuels nous permettent donc d'envisager la complexité de la gestion des processus rédactionnels. Nous nous intéresserons particulièrement au décours temporel de la production écrite reflétant cette gestion complexe, en posant notamment la question des unités de production.

Chapitre 3 - Étude du décours temporel de l'écriture

Le scripteur à qui il est demandé d'écrire un texte doit donc activer les processus rédactionnels tout en étant soumis à des contraintes, que ce soit d'ordre linguistique pour aboutir au texte fini souhaité, ou d'ordre cognitif puisque ses ressources sont limitées. Il est alors possible d'étudier le décours temporel de cette production, pour voir comment ces processus sont gérés.

Nous présenterons tout d'abord les outils à notre disposition pour étudier ces processus rédactionnels, en montrant l'évolution technique ayant permis d'ouvrir de nouvelles perspectives de recherche, et en indiquant les spécificités des différentes méthodes. Nous soulèverons ensuite la question qui sera au cœur de notre recherche : celle des unités de production, et nous exposerons les études ayant traité, comme nous souhaitons le faire, le lien entre les données chronométriques et le contenu linguistique. Puis nous montrerons en quoi ces études laissent de nombreuses questions ouvertes, et mentionnerons les difficultés d'ordre méthodologique qui se présenteront à nous avec ce paradigme d'étude. Enfin, nous présenterons notre question de recherche et la structure du travail élaboré ici.

1. Méthodes d'analyse de la gestion temporelle des processus rédactionnels

Différents protocoles ont été utilisés pour mettre à jour les processus rédactionnels (pour une revue de certaines méthodes, voir Fayol, 1997; Olive, 2002). Nous nous intéresserons plus particulièrement aux méthodes en temps réel, qui permettent d'étudier la gestion des processus dans leur aspect temporel ; contrairement aux méthodes a posteriori ou off-line, les indices sont alors recueillis en même temps que l'activité se déroule. Par ailleurs, nous mentionnerons les méthodes de recherche portant sur les retours sur le texte : l'objet de recherche que sont les traces de corrections et leur reconstruction temporelle, auquel s'ajoutent de nos jours de nouvelles techniques permettant de recueillir les données oculaires, peuvent en effet être également des indices quant à la gestion temporelle des processus rédactionnels.

1.1. Divers protocoles d'étude

L'étude des processus impliqués dans la production écrite pose des problèmes d'ordre méthodologique. En effet, la mise à jour des facteurs (empan de la MDT, fluence lors de la transcription, âge, connaissances encyclopédiques, etc.) ayant un impact sur le texte produit (qualité, longueur, complexité syntaxique, etc.) peut se faire en étudiant la corrélation entre le facteur observé et les caractéristiques du produit fini. En revanche, des protocoles différents doivent être utilisés si l'on souhaite comprendre les processus eux-mêmes, puisqu'il faut prendre des informations au cours de l'activité d'écriture. Le travail sur les processus rédactionnels peut avoir plusieurs objectifs : décrire les processus utilisés en fonction de certaines modalités, leur localisation, leur cout cognitif, etc. Un premier paradigme d'études a permis de mettre à jour les processus rédactionnels et continue à en étudier la gestion au cours de la rédaction et le cout cognitif de chaque processus : il s'agit de l'utilisation des protocoles verbaux et du temps de réaction. Un deuxième paradigme d'études utilise les données chronométriques recueillies pour déterminer les unités de production et les variations de charge cognitive en fonction de la tâche à effectuer. Enfin, des protocoles expérimentaux plus contraints permettent de répondre à des questions plus précises sur la gestion de ces processus, même si des limites quant à ce type de recherches sont à mentionner.

1.1.1. Les protocoles verbaux et la double tâche

Afin de mieux comprendre quelles ressources sont allouées à l'écriture, il est possible d'avoir recours à la méthode de la « double tâche ». Outre la tâche principale, dans notre cas l'écriture, le participant doit effectuer une autre tâche, comme mémoriser des mots, des chiffres, détecter un nouveau stimulus, effectuer des calculs (e.g. Olive, Kellogg & Piolat, 2008; Ransdell, Levy & Kellogg, 2002). Si les mêmes ressources cognitives sont partagées pour les deux tâches, les performances de la tâche principale ou secondaire se trouveront affectées, ce qui permet de mettre à jour la nature et la quantité

de ressources cognitives affectées à la tâche d'écriture (e.g. Kellogg, 1987; Olive & Kellogg, 2002). Cette double tâche a également été utilisée pour calculer le cout cognitif propre à une activité, en fonction de diverses contraintes ou localisations dans le texte : dans ce cas, il est demandé au scripteur de répondre à un signal sonore ou visuel présenté de manière relativement régulière (avec un délai de 15 à 45 secondes entre chaque stimulus) ; si le scripteur est engagé dans une activité plus couteuse, son « temps de réaction » sera plus long, ce qui permet de mettre à jour les variations de la mobilisation des ressources cognitives (Olive, 2004).

Les processus rédactionnels ont été mis à jour par Hayes et Flower (1980) grâce à la technique des protocoles verbaux : les scripteurs décrivent les activités mentales qu'ils effectuent en même temps qu'ils rédigent le texte. Il peut également arriver que cette verbalisation fasse l'objet d'un entretien après la composition du texte. Cette méthode permet en outre de voir à quels moments certains processus sont davantage mobilisés (Breetvelt et al., 1994). Associés à la technique de la « double tâche », les protocoles verbaux permettent de connaitre le cout de chaque processus. C'est ce que l'on nomme la « méthode de la triple tâche » : il est demandé au scripteur de composer un récit, de nommer le processus en cours et de réagir à un stimulus. Ainsi, le calcul du temps de réaction pour chaque type de processus permet d'en connaitre le cout cognitif (Olive, Kellogg & Piolat, 2002; Piolat & Olive, 2000). Il est également possible d'associer les verbalisations à l'activité motrice pour voir quel processus est associé de manière privilégiée avec les pauses ou les périodes d'exécution (Olive et al., 2009).

Les protocoles verbaux ont fait l'objet de diverses critiques. Ainsi, il y a eu des craintes quant au fait que le contenu des textes serait lésé par les fréquentes interruptions ou la verbalisation concomitante, ce qui est une critique qui peut également être adressée à la méthode de la double tâche. Il semble cependant que ces verbalisations n'aient pas d'impact sur la qualité du texte produit (Olive et al., 2002), mais que les processus puissent se trouver affectés en fonction du type de tâche demandé (Janssen, van Waes & van den Bergh, 1996), ce qui peut poser des questions quant à la fiabilité de certains résultats. Concernant la verbalisation rétroactive, la crainte est que le scripteur ait oublié les activités mentales effectuées. Enfin, de par leur nature, certains processus sont automatiques et donc pas forcément conscients, et ne seront par conséquent pas verbalisés par les scripteurs. Ces critiques n'invalident pas la pertinence des nombreuses études menées grâce aux protocoles verbaux, mais invitent à compléter les recherches portant sur les processus rédactionnels par d'autres méthodes. Notamment, l'activité de rédaction se déroulant dans le temps, le recours aux données chronométriques peut sembler judicieux pour mieux saisir la gestion temporelle des processus rédactionnels.

1.1.2. L'enregistrement des données chronométriques

Lors des premiers travaux, les données étaient collectées grâce à des enregistrements vidéos utilisant une caméra dotée d'un chronomètre (e.g. Fortier & Préfontaine, 1994; Matsuhashi, 1982). Cette méthodologie s'inspirait des recherches menées sur l'oral, pour lesquelles les données chronométriques

(durée des pauses remplies, des pauses silencieuses, de la fluence de production) étaient enregistrées (e.g. Boomer, 1965; Butterworth, 1975; Cook, Smith & Lalljee, 1974; Oliveira, 2002b). Les chercheurs établissaient ensuite un alignement entre les données recueillies et la rédaction, notant par exemple la localisation et la durée des pauses. Ce type de collectes, bien que très utile pour obtenir des premiers résultats, était limité quant à la précision des données recueillies (de l'ordre du dixième de seconde). Actuellement, divers dispositifs sont employés, suivant le mode d'écriture analysé. Pour les textes composés par ordinateur, trois logiciels sont notamment utilisés: Translog est particulièrement adapté à la collecte et l'analyse de traductions (Jakobsen, 2006; e.g. Immonen, 2006; Immonen & Mäkisalo, 2010) ; Inputlog enregistre toutes les saisies de l'ordinateur, tels que le traitement de texte, la consultation de mails ou de pages internet, et propose de nombreux outils pour l'analyse de données – indication de la durée des pauses, des révisions, segmentation automatique en mots et en phrases – (Leijten & van Waes, 2006; e.g. Leijten, van Waes & van Horenbeeck, 2015; Quinlan et al., 2012) ; enfin, Scriptlog propose un environnement plus contrôlé, permettant de présenter des images ou des amorces de texte lors des expérimentations (Strömqvist & Malmsten, 1998; e.g. Alves et al., 2007; Johansson, Johansson, Wengelin & Holmqvist, 2009). Pour collecter les données chronométriques lors de l'écriture manuscrite, il est possible de recourir à des tablettes à digitaliser associées à un logiciel d'enregistrement et traitement des données (e.g. Alamargot et al., 2014 ; Olive et al., 2009; van Hell et al., 2008). Les données enregistrées sont alors beaucoup plus précises (de l'ordre de la milliseconde).

Un avantage de cette méthode est qu'elle est peu intrusive, puisque le scripteur n'est pas interrompu dans son travail de rédaction. Certains outils, comme le logiciel Inputlog, permettent au participant d'utiliser son environnement de travail habituel. Les stylos utilisés avec les tablettes graphiques sont, dans la mesure du possible, similaires aux stylos à mine d'encre pouvant être utilisés quotidiennement.

La durée des pauses et la vitesse d'écriture sont considérées comme reflétant le cout des processus cognitifs : lorsque la charge cognitive est trop importante, le scripteur ralentit jusqu'à parfois s'arrêter (Schilperoord, 1996, 2002). La pause serait ainsi un indice de planification ou de révision : sa localisation associée à sa durée seraient donc des indices quant aux unités de production et au cout des processus engagés. Nous verrons cependant que de nombreux problèmes méthodologiques sont soulevés quant à l'utilisation des données chronométriques pour étudier les processus rédactionnels.

1.1.3. Le recours à des paradigmes expérimentaux fortement contrôlés

Les outils présentés jusqu'à présent sont généralement utilisés dans un cadre expérimental : on demande à des personnes de participer à une étude et on leur donne une consigne précise, qu'ils réalisent dans un environnement contrôlé. Même si des efforts sont fournis pour rendre cette situation la plus écologique possible (utilisation d'un stylo à mine d'encre, possibilité pour les enfants d'utiliser un brouillon comme en classe), il s'agit tout de même de cadres expérimentaux. Seuls certains logiciels,

comme Inputlog, peuvent être utilisés dans des situations tout à fait écologiques ; ce logiciel a notamment été utilisé pour étudier l'écriture de personnes dans leur cadre professionnel (Olive & Cislaru, 2015). Cependant, même s'il s'agit de situations plus ou moins expérimentales, les analyses sont souvent rendues difficiles car toutes les variables ne peuvent pas être contrôlées. Par exemple, dans une production de textes, la longueur des phrases va varier ; le lexique va être plus ou moins fréquent ; certains scripteurs vont faire plus ou moins de pauses. Cela peut limiter l'interprétation correcte de certains résultats. Par conséquent, pour éviter ces biais et répondre à une problématique bien précise, il est possible de recourir à des paradigmes expérimentaux fortement contrôlés. Par exemple, c'est ce qu'ont fait Quinlan et al. (2012) lorsqu'ils se sont intéressés de manière spécifique à la complétion de phrase et à la correction d'erreurs. Certains chercheurs recourent à des dénominations d'images pour faire produire des phrases (Torrance & Nottbusch, 2012), demandent à des scripteurs de compléter des phrases (Chanquoy et al., 1990) ou de produire une phrase précédemment entendue (Ferreira, 1991).

Des chercheurs ont combiné au cours de leurs recherches des approches expérimentales très contraintes, permettant de mettre en évidence des relations causales, et des études portant sur des corpus en utilisant des outils statistiques adéquats (corrélations, régressions), afin de généraliser davantage les résultats (Fayol & Lété, 2012). Cependant, il est important de rester conscient des limites de la généralisation de ces résultats. Tout d'abord, concernant les études expérimentales de manière générale, il faut souligner que le comportement du scripteur peut changer du fait qu'il participe à une étude. Cela peut être dû aux conditions expérimentales proposées qui diffèrent d'une situation d'écriture de la vie courante (support d'écriture différent, mouvements limités), à la présence d'un chercheur ou aux consignes imposées. Or, les modèles d'écriture ont bien souligné l'importance des composantes telles que la motivation, le destinataire ou le medium d'écriture ; ces éléments sont donc à prendre en compte dans l'analyse. Par ailleurs, les expériences dont les variables sont fortement contrôlées postulent qu'il est possible d'isoler une étape du processus pour en comprendre le fonctionnement, et que ce fonctionnement sera identique dans un cadre plus général. Or, des effets d'anticipation ou de retardement, tels que trouvés par Maggio, Lété, Chenu, Jisa et Fayol (2012), n'auraient par exemple pas pu être relevés dans une tâche de production de mot isolé.

1.2. Les retours sur le texte

1.2.1. L'étude des corrections

L'étude des corrections a fait l'objet de travaux en « critique génétique », ou « génétique textuelle », depuis les dernières décennies du XXème siècle (Lebrave, 2002). L'objet d'étude de ces chercheurs était la recomposition du cheminement de l'écriture, grâce à l'analyse des manuscrits. Fabre-Cols (2004) synthétise les principales caractéristiques de la critique génétique, afin de présenter l'apport de cette approche dans le domaine de la didactique de l'écrit. Dans le cadre de notre recherche, quelques points

retiendront particulièrement notre attention. Ainsi, les théories de l'énonciation (issues des travaux de Benveniste, Culioli, Fuchs), qui sont un ancrage théorique important dans cette discipline, invitent à percevoir à travers le texte final des textes virtuels, qui n'ont pas été conservés ; elles envisagent également le scripteur comme un lecteur, à travers le concept de « polyphonie énonciative », ce qui permet de compléter la façon dont est envisagée la lecture pendant l'écriture (Hayes, 1995).

Dans le domaine littéraire, cette approche a permis de considérer toute écriture comme un travail et une recherche, et de se détacher de la vision d'un texte final qui serait un ensemble clos. En effet, l'étude des manuscrits montre dans quelle mesure le scripteur a dû effectuer des choix parmi de nombreuses possibilités. Même s'il est parfois difficile, lors des analyses, de ne pas prendre en compte le texte final, il faut pourtant essayer d'analyser le texte tel qu'il est au point d'inscription. En outre, en s'intéressant de manière approfondie à des manuscrits d'un écrivain particulier, ce type de recherche permet d'en faire ressortir les caractéristiques d'écriture. Dans le domaine linguistique, le but est davantage de chercher ce qui peut être commun à l'ensemble des scripteurs, renseignant ainsi sur la production écrite de manière générale, ou d'essayer de déterminer des profils de scripteurs. Cela est certes possible à travers l'étude d'auteurs littéraires, lorsque les analyses détaillées portant sur certains auteurs sont comparées les unes aux autres, faisant émerger des processus communs à l'écriture littéraire. Cependant, une des particularités des linguistes est qu'ils ne s'appuient pas uniquement sur des corpus littéraires. Ainsi, des recherches utilisant les outils de la génétique textuelle ont porté sur des corpus d'élèves (e.g. Doquet-Lacoste, 2006; Fabre-Cols, 2004; Plane, 2003), afin de « saisir autant qu'il se peut, c'est-à-dire de façon lacunaire et problématique, les processus qui sont en œuvre dans l'écriture » (Fabre, 1991, p. 1). Dans le but de caractériser le processus d'écriture de manière complète, il est ainsi pertinent que des recherches portent à la fois sur des scripteurs en cours d'apprentissage, des adultes qui ne sont plus en situation d'apprentissage, et des adultes dont la profession requiert une expertise rédactionnelle (écrivains, journalistes, chercheurs).

Divers outils peuvent être utilisés pour étudier ces modifications du texte. Tout d'abord, il est possible d'observer le manuscrit, c'est-à-dire la feuille sur laquelle apparait la modification : il peut s'agir de manuscrits d'écrivains, mais également de brouillons ou d'écrits intermédiaires d'élèves collectés en classe, ou de textes produits dans des situations plus expérimentales, où le scripteur était autorisé à apporter des modifications sur son texte. De plus en plus de personnes, professionnels compris, utilisent désormais l'ordinateur pour écrire. Dès lors, le concept de « manuscrit » n'est plus pertinent, hormis si le scripteur imprime et annote certaines pages. En outre, le traitement de texte fait disparaitre de l'écran toute trace de modification du texte, rendant invisibles les traces de suppression ou d'ajouts (Doquet, 2006). Cependant, il est possible de recourir à des logiciels permettant d'enregistrer le déroulement de l'écriture et de le restituer, tels que Genèse du texte (produit par l'Association Française pour la lecture en 1993, e.g. Doquet, 1995, 2003a) ou Inputlog (Leijten & van Waes, 2006), la matériau disponible étant généralement appelé « écriture en temps réel » ou « écriture enregistrée »

(Doquet & Leblay, 2014). Ces logiciels permettent non seulement d'enregistrer toutes les opérations de modification, mais en indiquent en outre la temporalité : il est possible de savoir à quel moment la correction a été initiée, donc dans quel ordre les modifications ont été apportées, ainsi que la durée des pauses associées à ces modifications. De telles informations relatives à la temporalité des corrections sont aussi disponibles pour les textes manuscrits, par exemple grâce au logiciel Eye & Pen (Chesnet & Alamargot, 2005). Dans ces cas, l'écriture n'est plus reconstituée par les marques graphiques visibles sur le manuscrit, mais par la chronologie des opérations (Doquet, 2003). Grâce à ces outils, les retours sur le texte ayant laissé une trace graphique sont donc visibles. Cependant, ils ne permettent pas de mettre à jour toutes les fois où le scripteur a relu son texte sans que cette activité ne laisse de trace : c'est alors l'intérêt de l'enregistrement des mouvements oculaires de rendre disponibles ces traces de relecture.

1.2.2. Les mouvements oculaires

Les avancées techniques récentes ont permis de faire évoluer les outils dont les chercheurs disposent afin d'étudier la production écrite. Nous avons déjà mentionné l'existence de tablettes graphiques permettant de recueillir des données chronométriques avec une précision de la milliseconde, ou les logiciels qui enregistrent la composition de textes par ordinateur. Cependant, une des composantes de la production écrite reste jusque-là peu visible : il s'agit de la relecture. Or, les données relatives à la lecture du texte déjà écrit pourraient fournir des renseignements tant sur les processus de révision que de planification ou de formulation, et permettraient en outre d'expliquer la longueur de certaines pauses ou le ralentissement de la vitesse d'écriture.

Que ce soit lors de l'écriture à la main ou par ordinateur, l'utilisation d'un oculomètre permet d'enregistrer les mouvements des yeux. Certains logiciels, comme Eye & Pen (Alamargot, Chesnet, Dansac & Ros, 2006)**,** permettent de coupler l'enregistrement de données oculaires avec les données chronométriques, si la rédaction se fait à la main sur une tablette graphique. Des études portant sur la phrase utilisent les mouvements oculaires pour par exemple mieux comprendre des mécanismes comme l'accord du verbe avec le sujet (Alamargot et al., 2015), ou indiquer quand les images sont regardées dans le cas où une phrase est produite à partir d'images (Torrance & Nottbusch, 2012). Dans le cadre d'études portant sur le texte, il est par exemple possible de savoir de quelle manière les rédacteurs utilisent les ressources de leur environnement, qu'il s'agisse des consignes, des documents mis à disposition ou du texte en cours d'élaboration, et d'étudier la manière dont se coordonnent la lecture et l'écriture (Alamargot, 2005; Alamargot, Dansac, Ros & Chuy, 2005). Une étude récente a en outre montré que les séquences de relecture du texte n'étaient pas initiées de manière aléatoire : elles sont plus fréquentes et comportent davantage de fixation en frontière de phrase qu'aux frontières de plus bas-niveau (Torrance, Johansson, Johansson & Wengelin, 2015), suggérant que la lecture puisse jouer un rôle dans la planification de la phrase suivante.

Une des difficultés de l'analyse des relectures est que celles-ci peuvent avoir plusieurs buts, dont le fonctionnement n'est pas le même. Hayes (1995) avaient déjà distingué deux modèles de lecture différents : lire pour comprendre et lire pour réviser. Or, « lire pour comprendre » est surtout utilisé dans le cadre d'une tâche de lecture. Certes, l'analyse des mouvements oculaires a été utilisée avec succès pour comprendre les mécanismes de la lecture, que ce soit au niveau du mot (e.g. Clifton, Staub & Rayner, 2007 pour une revue) ou du discours (e.g. Baccino & Pynte, 1998; Hyönä, Lorch & Rinck, 2003; Just & Carpenter, 1980). Cependant, les résultats obtenus en lecture d'un texte ne sont pas nécessairement applicables lorsque la lecture se fait sur un texte qui est en train d'être écrit : Johansson, Johansson, Wengelin et Holmqvist (2012) ont par exemple montré que, dans ce dernier cas, il y avait davantage de retours oculaires sur la ligne précédente, sans doute pour des nécessités de cohérence ou pour réviser le texte précédemment écrit. Pendant une tâche d'écriture, on pourrait en outre ajouter l'activité « lire pour poursuivre le texte ». Dans ce type de lecture, le but n'est pas d'identifier un écart entre sa production et la représentation mentale qui en était faite ; il s'agit en fait de s'appuyer sur le texte déjà écrit pour le continuer de manière cohérente, et en utilisant les outils de cohésion adéquats. Même si les processus rédactionnels seront sans doute mieux compris avec l'utilisation des mouvements oculaires, ces difficultés d'interprétation devront être prises en compte lors des analyses.

Les diverses études utilisant les mouvements oculaires montrent combien cette technique ouvre de nouvelles perspectives dans le champ de la recherche en production écrite, et invitent à davantage utiliser ces outils dans le cadre de l'écriture de textes. Malheureusement, certaines limites techniques empêchent parfois de recueillir des données oculaires très précises sur des gros corpus, ou avec différents types de participants : il faut en effet s'assurer que le participant ne bouge pas trop la tête pour éviter le dérèglement des paramètres, et vérifier que les données oculaires recueillies sont bien alignées avec l'enregistrement du texte écrit.

1.3. Conclusion sur les méthodes à notre disposition

Ainsi, les chercheurs ont à leur disposition de nombreux outils pour étudier les processus rédactionnels. Ces méthodes sont issues de divers champs disciplinaires : nous avons mentionné aussi bien des protocoles expérimentaux fortement contraints, caractéristiques des études menées en psychologie cognitive, que des méthodes d'analyse de textes issues de la critique génétique. L'étendue des données analysées varie également : il peut s'agir de quelques phrases produites par des dizaines de participants, comme d'un unique texte (mais constituant parfois un livre) produit par quelques personnes. Certaines études ont vocation à informer sur un processus particulier, en considérant que sa gestion sera identique chez tous les participants d'un groupe constitué comme « homogène », alors que d'autres recherches s'intéresseront aux variabilités interindividuelles, en cherchant parfois tout de même à établir des « profils ».

Chacune de ces méthodes possède ses limites, mais nous sommes convaincus que la complémentarité de ces approches peut fournir de précieux renseignements quant à la gestion temporelle des processus rédactionnels. De plus, l'utilisation de certaines informations, comme la chronologie absolue et relative des révisions, peut éclairer des résultats obtenus grâce aux paradigmes plus classiques que sont par exemple les analyses portant sur les pauses.

Nous allons à présent introduire la question qui guidera notre recherche, à savoir la taille des unités de micro-planification et de formulation et la gestion temporelle de leur production. La temporalité de l'écriture sera étudiée dans nos travaux à travers la mise en relation des données chronométriques (durées des pauses et des temps d'exécution[23]) avec le contenu linguistique du texte, ainsi que par l'étude de la temporalité des révisions.

[23] La notion de « temps d'exécution » correspond dans nos travaux au temps mis par le scripteur pour écrire une unité, une fois toutes les pauses (supérieures à 15 ms, qu'elles soient inter- ou intra-mots) exclues. Cela rejoint la notion de « débit » de certains auteurs (e.g. Lété, Maggio). La chronométrie associée à l'exécution est parfois appréhendée à travers la notion de « vitesse d'écriture » ou de « périodes d'exécution », ce dernier terme étant généralement réservé aux études qui délimitent un seuil pour les pauses afin d'exclure les pauses « non cognitives », et identifiant donc une « période d'exécution » à une période située entre deux pauses, durant laquelle au moins un mot est écrit.

2. L'identification des unités de production grâce aux données chronométriques

2.1. La question des unités de production

Les modèles présentés précédemment, ainsi que les discussions concernant la gestion des processus cognitifs, amènent la question des unités de production. Si l'on prend le modèle de Levelt (1999), on parle de le construction d'un message pré-linguistique, de son encodage et de son articulation, mais il n'est pas précisé à quelle « unité » de la langue ce modèle s'applique. Levelt (1989) soulignait déjà cette question, en rappelant que différentes unités ont été proposées, sur la base de preuves empiriques comme la durée des pauses, les erreurs de production, la prosodie ou les gestes accompagnant le langage : idées, syntagmes, clauses « profondes » ou « de surface », phrases, etc. Devant la multitude des études menées et des preuves apportées, il parait évident, comme le souligne Levelt, qu'« il n'y a pas une unique unité de parole » (1989, p. 23). Il conviendra cependant de voir si certains consensus émergent de ces différentes études.

On peut par ailleurs remarquer que, si Olive (2014) ne formule pas de proposition quant à l'unité « n » qui est conceptualisée, formulée puis transcrite, la taille de cette unité présentée dans son schéma (Figure 6) varie, ce qui suggère l'existence de plusieurs unités, de taille différentes. Hayes et Flower (1980), par exemple, proposent que l'unité gardée en mémoire est la proposition, entendue dans un sens sémantique, par exemple une structure exprimant une relation entre deux concepts, ou appliquant un attribut à un concept. Ces propositions seraient récupérées en mémoire lors de la planification de la phrase, puis la phrase serait écrite proposition par proposition ; une fois la phrase finie, la phrase suivante serait planifiée, puis transcrite. Lebrave (2009) suggère que la formulation ne dépasse pas la phrase en train d'être écrite, la suite n'étant anticipée que de manière vague, sous forme « d'intention de communication », qui pourrait correspondre à ce qui a été macro-planifié.

Van Galen (1991), dans son modèle qui décrit principalement l'étape de l'exécution motrice, propose que les processus ne concernent pas des unités de même taille : l'activation des intentions porte sur les idées, la récupération sémantique sur les concepts, la construction syntaxique sur les phrases, l'orthographe sur les mots, la sélection des allographes sur les graphèmes, le contrôle de la taille sur les allographes, et l'ajustement musculaire sur les traits. Cette proposition est sans doute schématique, mais elle souligne le fait que le scripteur est amené à prendre des décisions sur des unités de plus en plus petites. Concernant les unités de « haut-niveau », notamment de micro-planification et formulation, on peut s'interroger sur l'empan de la réflexion[24]. Par exemple, « l'idée » consiste-t-elle en une clause, ou

[24] De manière rigoureuse quant aux définitions, nous préférons le terme d'unité de « réflexion » à celui d'unité de « planification », car à l'échelle d'une phrase la réflexion peut porter aussi bien sur la micro-structure que sur la

une phrase ? Si chaque idée correspond à une clause, les différentes idées constituant une même phrase sont-elles planifiées en même temps, ou chaque clause équivalant à une idée est-elle planifiée de manière indépendante ? En outre, si les idées sont formulées les unes après les autres, le scripteur attend-il la fin de la transcription d'une idée pour réfléchir à l'idée suivante, ou la planification d'une idée peut-elle être initiée pendant la transcription de l'idée précédente ? Ces problématiques sont résumées par Plane, Alamargot et al. (2010, p. 17) :

> « Ce mode de traitement permet de concevoir plusieurs temporalités simultanées au cours de l'activité, associées à des unités et traitements différents. Ce fonctionnement en strates temporelles – caractérisées par des différences de rythme, d'empan et d'unités – explique que la production d'écrit est soumise à des rétroactions et des anticipations, lesquelles s'exercent sur des objets d'échelles différentes, allant du fragment de texte à la simple unité linguistique, et avec des « grains de traitements ou de représentations » différents. »

Par ailleurs, ces unités de traitement peuvent différer en fonction du niveau d'expertise du scripteur. En effet, nous avons soulevé le problème de la segmentation des textes pour les enfants et adolescents. Or, si les jeunes scripteurs ont du mal à appréhender la notion de « phrase », il est sans doute peu pertinent de supposer que cette unité soit une unité de planification. De même, la forte proportion de clauses juxtaposées ou coordonnées dans leur texte (Mazur-Palandre, 2007), reflétant vraisemblablement la stratégie de « connaissances rapportées », peut sans doute amener les scripteurs à planifier leur texte sur des unités plus petites telles que la clause. Enfin, la question de l'empan de planification étant liée à celle de la mémoire, il est possible que les plus jeunes ne soient pas en capacité de planifier leur texte sur une grande unité, s'ils ne peuvent pas ensuite la garder en mémoire le temps de la transcription.

Ainsi, nous sommes amenés à nous interroger sur l'unité de production et sur la gestion temporelle des processus. Pour cela, nous avons privilégié l'analyse des données chronométriques enregistrées lors de la production du texte, en les mettant en lien avec le contenu linguistique. Nous allons à présent détailler les principales études ayant cherché à déterminer l'empan de planification lors de la production du langage.

2.2. Lien entre données chronométriques et contenu linguistique

Les pauses étant généralement associées à la planification de ce qui suit (Matsuhashi, 1982), de nombreuses études ont cherché à déterminer quelles étaient les unités de production en mettant en lien

---

formulation. Lorsque nous utiliserons le terme « d'unité de planification », terme employé dans de nombreuses études, ce sera dans un sens identique, à savoir « ce qui est planifié », cette planification pouvant porter sur les aspects sémantiques, la structure syntaxique ou le lexique. Le terme « unité de production », parfois utilisé, souligne quant à lui que la taille de l'unité de réflexion peut être appréhendé par l'étude de la dynamique d'écriture, en observant par exemple où sont les pauses principales, ou à quel endroit le scripteur effectue un retour sur son texte.

les données chronométriques et le contenu linguistique de la production. Dans ce cadre, il est considéré que les variations de durée de pause et de débit d'écriture sont dues à des variations de cout cognitif : plus la charge cognitive est importante, plus la pause sera longue et le débit sera lent (Fayol & Lété, 2012). Ces études diffèrent fortement l'une de l'autre au niveau méthodologique, que ce soit concernant la modalité (oral ou écrit), le type de protocole (production spontanée, histoire à imaginer, dénomination d'images) ou l'étendue de la production (texte, phrase, syntagme). Cependant, certains résultats sont récurrents à travers les études, et peuvent donc servir de base solide pour la suite des travaux. En outre, les différents résultats obtenus, même s'ils ne seront pas répliqués dans des conditions similaires, peuvent ouvrir des pistes d'explorations intéressantes. Nous présenterons tout d'abord les études expérimentales portant sur la phrase ; puis nous aborderons la question des unités de production au sein de discours ou de textes[25], et enfin d'autres facteurs pouvant expliquer les variations de durée des pauses seront évoqués.

2.2.1. Production d'une phrase

2.2.1.1. À l'oral

Les premières études utilisant les données chronométriques, que ce soit au niveau de la phrase ou du discours, portaient sur l'oral. Certes, des différences existent entre la production orale et la production écrite. Cependant, il est possible que certains processus de production soient caractéristiques du langage de manière générale, et que certains effets soient donc aussi bien présents à l'écrit qu'à l'oral ; c'est pourquoi il demeure intéressant de connaitre les résultats de certaines études portant sur l'oral.

Des études de psychologie cognitive dont le protocole expérimental était très contraint ont abordé la question de l'empan de planification lors de la production d'une phrase. Il s'agissait généralement de décrire une série d'images, avec soit une variation de la structure de la phrase, soit des variations quant au degré d'accessibilité du contenu lexical ou sémantique d'un des termes. Suivant les effets de ces variations sur la durée de la latence initiale ou à d'autres emplacements de la phrase, les auteurs proposaient un empan de planification.

Tout d'abord, Griffin (2001, 2003) propose que la production soit essentiellement incrémentale. En effet, dans une étude où des locuteurs anglophones devaient produire des phrases du type « A et B sont au-dessus de C », la facilité de retrouver les termes B et C (suivant leur fréquence ou le nombre de dénominations possibles) n'affectait pas la latence initiale. En outre, dans le cas de sujets complexes de type « A et B », la latence initiale était plus longue uniquement si « A » était monosyllabique et qu'il y avait peu de mots entre A et B. Cela montrerait que, de manière privilégiée, les locuteurs préparent la production des items au fur et à mesure de la production de la phrase.

[25] Ces deux termes renvoient à une production monologique : « discours » est employé pour une production orale, et « texte » pour une production écrite.

Cependant, de nombreuses études suggèrent plutôt que la planification initiale porte au-delà du premier groupe nominal. Smith et Wheeldon (1999) proposent que l'empan de planification soit le premier argument du verbe, i.e. le sujet, puisque la latence initiale est plus longue si le sujet est complexe (de type « A et B sont au-dessus de C ») que s'il est simple (« A est au-dessus de B et C »). De même, Levelt et Maassen (1981) montrent que la latence initiale est plus longue dans des phrases avec un sujet complexe (de type « A et B montent ») que s'il y a deux clauses coordonnées (« A monte et B monte »). Cependant, la position du syntagme peut être plus importante que sa fonction syntaxique, puisque Allum et Wheeldon (2007, 2009) ont montré, en se basant sur des productions en anglais et en japonais, que la planification portait toujours sur le premier syntagme, même quand celui-ci ne correspondait pas à la tête syntaxique du sujet. Certains auteurs suggèrent un empan plus important : par exemple, dans le cas d'une dénomination de deux images où un distracteur sémantique est présenté, il y a un effet d'interférence conduisant à une latence initiale plus longue, que le distracteur soit lié à l'objet A ou B, ce qui signifie que l'accès sémantique a été complété pour la clause entière avant le début de la production (Meyer, 1996). Cependant, Schriefers, Teruel et Meinshausen (1998) proposent que le verbe n'est pas toujours planifié dès le début. En effet, en faisant produire en allemand des phrases dont le verbe était soit en position initiale, soit en position finale, ils n'ont trouvé un effet d'interférence avec un verbe relié sémantiquement que lorsque le verbe était en début de phrase. Cependant, la structure des clauses avec le verbe en position finale était souvent déductible de l'aspect animé ou non animé du sujet, ce qui pouvait expliquer qu'il n'était pas nécessaire pour les locuteurs de traiter le prédicat dès le début de la phrase pour la construire correctement.

Par ailleurs, Ferreira (1991) s'est intéressée à la production de phrases dans le cadre d'une tâche de répétition, donc pour laquelle il n'y avait pas de processus de micro-planification ni de formulation nécessaire, mais où les locuteurs devaient générer des représentations phonologiques, phonétiques et motrices. Un effet de longueur sur la latence initiale a été observé, les phrases les plus longues étant précédées d'une pause plus longue, ce qui suggère que la planification dépasse le simple mot. En faisant varier la complexité des syntagmes, il apparait que la présence d'un sujet complexe augmente la durée de la pause initiale, et qu'il y a une plus longue pause avant le verbe si l'objet est complexe, la phrase étant donc sans doute produite en deux parties lorsqu'elle est plus complexe.

Il y a donc un consensus sur le fait que les locuteurs planifient généralement au moins le sujet de la phrase, même s'ils ne planifient pas la clause entière. Comme le soulignent Martin, Crowther, Knight, Tamborello et Yang (2010), une planification plus incrémentale est possible lorsqu'il y a des contraintes de fluence liées à la modalité orale, ou lorsque la répétition d'une même structure syntaxique au long de l'épreuve implique une simple planification lexicale.

2.2.1.2. À l'écrit

Certaines études sur la production écrite de mots ou de syntagmes nominaux ont utilisé des paradigmes expérimentaux permettant le contrôle rigoureux des multiples facteurs pouvant affecter la production, permettant de reporter par exemple des effets d'âge d'acquisition (Bonin, Fayol & Chalard, 2001) ou de phonologie (Zhang & Damian, 2010). Récemment, des études portant sur la phrase écrite ont tenté de répliquer les résultats obtenus sur la production de phrase orale. Ainsi, Roeser, Torrance et Baguley (2015) ont répliqué l'étude de Smith & Wheeldon (1999), en faisant varier la complexité du sujet et la facilité d'accès lexical du nom « B ». Il est apparu que la latence initiale est également plus longue avant un sujet complexe qu'un sujet simple, et que l'effet de codabilité[26] apparait dans le cas d'un sujet complexe, mais qu'il apparait de manière plus marquée avant la production du nom B. L'empan semble donc également être le sujet à l'écrit, même si l'accès lexical n'est sans doute pas achevé avant le début de la production. Dans une tâche de dénomination de deux à quatre images, Torrance & Nottbusch (2012) ont également fait apparaitre que l'empan pouvait être au minimum le sujet, puisque les phrases « complexe-complexe » (« A et B sont au-dessus de C et D ») étaient initiées par une pause plus longue que les phrases « simple-complexe » (« A est au-dessus de B et C ») et « complexe-simple » (« A et B sont au-dessus de C »), les pauses initiales les plus courtes se situant avant les phrases « simple-simple » (« A est au-dessus de B »). Comme pour l'oral, les résultats ne sont pas tous concordants. Ainsi, en reproduisant l'expérience de Griffin (2001), dans laquelle les scripteurs devaient décrire des images sous la forme « A et B sont au-dessus de C », « B » étant plus ou moins facile d'accès, Torrance et White (cités dans Torrance & Nottbusch, 2012) n'ont pas trouvé d'effet lié à la facilité d'accès de « B », que le sujet soit coordonné ou subordonné. En outre, le cout de la planification peut également dépendre des liens de dépendance entre les constituants. Ainsi, les locuteurs allemands ont des latences plus longues avant des phrases dont le sujet complexe est constitué d'un nom « B » subordonné au nom « A » (avec la préposition « avec ») que lorsque les deux noms sont coordonnés. Dans le cas d'une coordination, il y a une pause plus longue avant de produire la deuxième partie du sujet.

En demandant aux participants d'écrire la fin d'une histoire, Chanquoy et a1. (1990) ont étudié dans quelle mesure les données chronométriques variaient en fonction de la complexité de la phrase. Les scripteurs devaient relater trois actions ou trois états, correspondant donc à trois clauses, soit en les regroupant au sein d'une même phrase, soit en faisant trois phrases différentes. Il est apparu que le temps de pré-écriture était plus long si le participant devait composer une seule phrase ; en outre, les pauses avant la clause 2 et avant la clause 3 étaient de durée équivalente au sein d'une même phrase, mais la pause avant la dernière clause était plus longue que la précédente s'il s'agissait de trois phrases. De manière générale, la vitesse d'écriture décroissait d'une clause à l'autre. Cette étude suggère donc un

[26] c'est-à-dire le nombre de mots disponibles pour décrire une image

empan de planification assez large des trois propositions, avec une stratégie plus incrémentale lorsqu'il s'agit de phrases distinctes.

Tout comme les études portant sur l'oral, les études portant sur la production écrite de phrases ne décrivent pas toutes le même empan de planification, mais suggèrent que celle-ci porte au-delà du premier syntagme. Elles soulignent en outre que des variations quant à cet empan peuvent être dues au lien de dépendance entre les constituants, c'est-à-dire à la structure syntaxique de la phrase.

2.2.2. Production d'un texte

L'étude de la production d'un texte écrit ou d'un discours oral est délicate car il n'est pas possible de contrôler toutes les caractéristiques des mots et des phrases, comme le permettent les paradigmes expérimentaux décrits dans le cadre des études sur la phrase. Les productions étudiées ont cependant l'avantage d'être globalement moins contraintes, reflétant ainsi davantage la manière dont un monologue long est produit. En effet, rien ne garantit que les processus engagés lors de la production de phrases isolées se déroulent de manière identique dans le cas d'un texte. La question principale guidant de nombreuses recherches qui portent sur des textes est : quelles sont les unités fonctionnelles de planification (Matsuhashi, 1982) ? Là encore, le recours aux données chronométriques a permis de fournir de nombreux indices, puisque la longueur de la pause précédant une unité était considérée comme reflétant la planification de ce qui suivait : plus la pause était longue, plus la planification était couteuse, en termes de longueur ou de contenu à planifier. Pour explorer la question des grandes unités de planification, les chercheurs segmentaient le texte en unités sémantiques ou syntaxiques, et comparaient la durée des pauses ou la fréquence d'apparition des pauses aux différents niveaux hiérarchiques[27].

2.2.2.1. Structure du texte

Au niveau de la macrostructure, certains chercheurs ont essayé d'établir un lien entre la structure du texte et la longueur des pauses. Ainsi, Schilperoord et Sanders (1999) ont fait ressortir la structure des relations hiérarchiques du texte grâce à une analyse PISA[28] : la longueur des pauses reflétait bien la hiérarchie de la structure, les pauses les plus courtes étant celles à l'intérieur des segments, alors que les pauses les plus longues précédaient des changements au niveau de la ligne principale. Après avoir fait produire des textes narratifs à l'oral, Oliveira (2002b) a demandé à des experts en linguistique discursive de segmenter le texte à partir du modèle de Labov, puis a mis en relation les pauses silencieuses

[27] Les « niveaux hiérarchiques » désignent les frontières d'unités linguistiques : paragraphe, phrase, clause, syntagme…

[28] Acronyme pour Procedures for Incremental Structure Analysis : procédure permettant de fournir une description structurale des textes, notamment concernant les aspects liés aux liens sémantiques (e.g. problème - solution) et à la nature hiérarchique des connections (segments coordonnés ou subordonnés).

supérieures à 250 ms avec les frontières narratives. Il ressort que la majorité des frontières narratives (65%) sont marquées par des pauses, ce qui masque des variations interindividuelles, puisque certains locuteurs font des pauses à toutes ces frontières narratives alors que certains n'en font jamais, ce profil étant plus rare. En comparant les frontières narratives avec les autres frontières[29], il ressort que les frontières narratives sont en moyenne plus longues. La fréquence et la durée de ces pauses ne dépendent pas de la section narrative, même si les pauses présentes dans la coda tendent à être moins fréquentes mais plus longues (Oliveira, 2002a).

Sans recourir à une segmentation se basant sur un modèle bien défini, Butterworth (1975) a segmenté les textes argumentatifs oraux recueillis en grandes « unités sémantiques », qui peuvent correspondre à une phrase ou plus ; la « phrase » est sans doute définie en terme de complétude syntaxique, puisqu'il s'agit de production orale. Sans que des données chiffrées indiquent dans quelle mesure les frontières d'unités sémantiques correspondent à des frontières de cycle, l'auteur indique qu'il y a une bonne adéquation entre phrases, idées et cycles, et suggère de prendre en compte la dimension sémantique lors de la segmentation. Il pourrait être intéressant de savoir dans quelle mesure ces « unités sémantiques » correspondent aux « phrases graphiques » de l'écrit. Hormis les pauses, il est possible de s'intéresser également à la fluence, et il apparait que celle-ci diminue lorsqu'il y a un changement de grande idée, dans le cadre de discours argumentatifs oraux par exemple (Greene & Cappella, 1986), sauf si la structure générale a été donnée avant le début de la production.

2.2.2.2. Emplacement syntaxique

Certaines unités de segmentation ne sont pas pertinentes à l'oral, car leurs définitions se basent sur des critères typographiques : c'est par exemple le cas du paragraphe, qui se distingue par un retour à la ligne et par la présence d'un alinéa, ou de la phrase définie avec des critères typographiques (débutant par une majuscule et / ou finissant par un point). Les unités choisies à l'oral par les linguistes peuvent en revanche s'appuyer sur des critères syntaxiques et prosodiques. Ainsi, Boomer (1965) s'est intéressé à la position des pauses (supérieures à 200 ms) au sein de la « clause phonémique », définie comme unité prosodique minimale. Il note que les hésitations sont essentiellement situées entre le premier et le deuxième mot de la clause phonémique, suggérant que l'encodage du discours se fait effectivement au-delà de l'unité « mot ». En étudiant les pauses silencieuses dans des discours spontanés ainsi que dans des tâches de lecture, Goldman et al. (2010) ont identifié que 94% des pauses longues se situaient en frontière « d'unité de rection » (correspondant sans doute à notre définition de la « clause »), alors que c'était le cas de seulement 17% des pauses courtes ; en outre, les locuteurs utilisaient fréquemment la pause longue pour reprendre leur souffle. Utilisant une segmentation plus détaillée, Piolat (1983) distingue les pauses à l'intérieur des clauses, les pauses entre clauses coordonnées ou subordonnées et

[29] Il n'est pas précisé ce qu'est une « frontière non narrative », mais on peut supposer qu'il s'agit d'une frontière syntaxique.

les pauses entre clauses juxtaposées, ce dernier niveau correspondant à ce que nous avons appelé « unité T ». En prenant en compte les pauses supérieures à 300 ms, il est apparu que les pauses étaient en moyenne plus longues entre les clauses juxtaposées qu'entre les autres clauses, les pauses internes aux clauses étant les plus courtes. Cependant, les résultats basés sur les moyennes peuvent masquer des variations au cours du texte, les frontières les plus longues ne correspondant pas de manière systématique aux frontières syntaxiques les plus élevées. Ainsi, la mise en relation des critères prosodiques et syntaxiques a permis à Degand et Simon (2009) de faire ressortir des « unités discursives de base », délimitées par une frontière à la fois syntaxique et prosodique ; il pouvait arriver que ces unités discursives de base soient composées d'une seule unité syntaxique mais de plusieurs unités prosodiques, d'une seule unité prosodique mais de plusieurs unités syntaxiques, ou qu'il y ait une adéquation entre unité syntaxique et unité prosodique.

À l'écrit, l'effet de l'emplacement syntaxique a été vérifié de nombreuses fois : les pauses avant les paragraphes sont en moyenne plus longues que celles avant les phrases, elles-mêmes plus longues que celles avant les clauses à l'intérieur des phrases, les pauses à l'intérieur des clauses étant les plus courtes. Cela a été vérifié dans plusieurs langues, comme le français (Chanquoy, Foulin & Fayol, 1996; Foulin, 1998), l'anglais (Sanders & Schilperoord, 2006; Schilperoord, 1996; Schilperoord & Sanders, 1999), le finnois (Immonen, 2006) ou le néerlandais (van Hell et a1., 2008), et ce dans différents types de textes (descriptif, argumentatif, expositif, narratif). Certains résultats diffèrent légèrement : par exemple, la pause avant le paragraphe n'est pas toujours plus longue que celle avant les autres phrases (Foulin, 1998), ce qui est peut-être dû à une grande variabilité et un nombre peu important de données. Il faut en outre mentionner que la définition même de l'emplacement syntaxique peut porter à discussion. Ainsi, alors que la plupart des chercheurs s'appuient sur le texte final pour segmenter la production, Spelman-Miller (2002a, 2002b) segmente uniquement par rapport au texte déjà écrit, indiquant les frontières par rapport au point de complétion potentiel : par exemple, si à un certain endroit la clause pourrait être terminée, il est indiqué que c'est une frontière de clause. Ce type de définition nécessiterait de préciser s'il s'agit d'une complétion syntaxique, sémantique ou communicative, l'établissement d'une frontière pouvant être délicat dans ces derniers cas.

2.2.2.3. Type de clause

Quelques études sont allées au-delà de la segmentation en différentes unités syntaxiques, et ont observé plus précisément le type de clause qui était produit. L'étude du type de clause est également liée à la question de l'unité de planification : ainsi, il est fait l'hypothèse que les clauses les plus dépendantes seront planifiées de manière anticipée. Au préalable, les résultats concernant le lien entre longueur des pauses et emplacement syntaxiques étaient toujours confirmés. Dans des corpus d'oral spontané (Goldman-Eisler, 1972), 78% des pauses entre phrases duraient plus de 500 ms ; 66% duraient entre 250 et 500 ms, alors que seuls 7% des pauses à l'intérieur des clauses duraient plus de 250 ms. De manière plus précise, la quantité de pauses supérieures à 500 ms était plus grande avant une clause coordonnée

qu'avant une clause subordonnée, le taux étant le plus faible pour les relatives. Des pauses de démarcation restaient présentes en lecture, mais les pauses étaient globalement moins longues, avec particulièrement la disparition des valeurs extrêmes pour les phrases. Les clauses subordonnées de textes écrits par ordinateur par des Finnois en langue maternelle (Immonen & Mäkisalo, 2010) étaient quant à elles précédées d'une pause plus courte que les clauses principales ; les pauses à l'intérieur de la clause principale étaient en outre plus longues que celles à l'intérieur de la subordonnée, les auteurs suggérant que la subordonnée est planifiée dans une certaine mesure pendant l'écriture de la clause principale. Piolat (1983) a identifié différents types de clauses, qui dans l'ordre décroissant de la durée de la pause les précédant étaient : coordonnées ; relatives, complétives, interrogatives indirectes ; circonstancielles. Ces résultats n'étaient pas tout à fait corroborés par van Hell et al. (2008), qui avaient remarqué que les pauses précédant les clauses principales étaient plus longues que les autres, mais qu'il n'y avait pas de différences entre les clauses circonstancielles, subordonnées non finies et relatives, ni entre relatives restrictives et non-restrictives. À l'oral, quelques catégories particulières ont été étudiées. Ainsi, la quantité de pauses supérieures à 200 ms était inférieure devant les clauses ayant un verbe non fini, lorsque des locuteurs anglais devaient poursuivre une histoire à l'oral (Holmes, 1984). Rochester et Gill (1973), quant à eux, n'ont pas relevé davantage de perturbations en fonction de la complexité de la phrase, c'est-à-dire du nombre de subordonnées, lorsque la taille était contrôlée ; en revanche, il y avait davantage de perturbations lors de la production d'une subordonnée complétive complétant un nom (du type « le fait que… ») que lors de la production d'une subordonnée relative (du type « le livre qui… »).

Ces résultats, portant sur des textes, suggèrent que la phrase est un tout, une réalité psychologique. Il y aurait particulièrement une planification globale lorsque les clauses dépendraient davantage les unes des autres : dans le cas des subordonnées, notamment lorsqu'elles sont enchâssées dans la principale. Cela n'empêche pas que des clauses puissent être ajoutées, surtout lorsqu'elles ne sont pas fortement liées, d'un point de vue sémantique et/ou syntaxique, aux clauses précédentes.

2.2.3. Facteurs affectant le rythme d'écriture des unités de production

2.2.3.1. Facteurs linguistiques

La majorité des études ont donc pour but de déterminer les grandes unités de planification. À partir de là, il est possible de s'interroger sur les facteurs qui influent les variations des données chronométriques.

Un des critères de variation peut être le type de texte ou de situation. Ainsi, lors de la rédaction de la fin d'un texte, les pauses entre les clauses tendent à être plus longues pour les textes narratifs que descriptifs (Chanquoy et al., 1990). À l'oral, 99% des pauses sont inférieures à 2 secondes lors de discussions, alors que la quantité de pauses supérieures à 3 secondes est respectivement de 5,4% et

10,4% pour des descriptions et des résumés (Goldman-Eisler, 1961). Dans leur étude sur les unités discursives de base, Degand et Simon (2009) ont montré que les unités syntaxiques regroupant plusieurs unités prosodiques étaient rares dans le cadre de conversations, mais plus fréquentes lors d'interviews ou d'informations à la radio. Dans ce cas, nous constatons que les contraintes communicatives inhérentes à un dialogue réduisent la proportion de pauses longues, alors que celles-ci peuvent être plus fréquentes dans d'autres contextes, à des fins de planification ou de perception à l'intention de l'auditeur, lorsque certains éléments sont mis en valeur par la prosodie. En outre, les scripteurs en situation de traduction semblent adopter une stratégie différente que lors de l'écriture d'un monologue dans leur langue maternelle, avec des pauses initiales moins longues, une planification davantage clause par clause, et plus de pauses à l'intérieur des clauses. En effet, lors de traductions anglais - finnois par des traducteurs professionnels finnois (Immonen, 2006; Immonen & Mäkisalo, 2010), il a été relevé que les pauses aux emplacements hiérarchiques les plus élevés, i.e. paragraphes et phrases, étaient moins longues lors de la traduction, et qu'il n'y avait pas de différence entre durée des pauses avant une clause subordonnée ou avant une clause principale. De plus, en traduction les pauses à l'intérieur des clauses sont en moyenne plus longues que dans le texte en finnois.

Un effet de longueur avait été soulevé par Cook et al. (1974) dans des interviews, puisque les clauses suivant une pause remplie étaient en moyenne plus longues que celles qui n'étaient pas initiées par une pause remplie. Ils n'avaient en revanche pas trouvé de corrélation générale entre la complexité syntaxique des phrases et le nombre de pauses remplies, ces pauses reflétant sans doute davantage une hésitation au niveau de la clause. En outre, les amorces syntaxiques ou la facilité d'accès au lexique peuvent aussi faire varier la durée des pauses, dans le cadre de production de phrases orales (Konopka, 2012; Konopka & Meyer, 2014).

Les clauses peuvent également varier quant à leur contenu sémantique, notamment le nombre et la nature des nouveaux cas introduits, ou suivant la complexité morpho-syntaxique (Bertrand & Dumoulin, 2013). Ainsi, dans des textes produits par des élèves de 5ème et de 3ème, les pauses initiales des clauses sont plus longues lorsqu'il y a davantage de cas nouveaux et une complexité morphosyntaxique plus élevée, et le débit est plus lent s'il y a davantage de cas nouveaux. Lorsqu'un seul cas nouveau est introduit, la pause initiale a tendance à être plus longue lorsqu'il s'agit d'une information temporelle qu'une information concernant le lieu ; le débit est quant à lui affecté par le type de personnage introduit, puisqu'il est plus lent s'il y a une introduction d'un bénéficiaire plutôt que d'un agent. Concernant l'aspect prévisible du contenu de la clause, Chanquoy et al. (1990) ont montré que plus la fin était prévisible, plus la pause de pré-écriture et les pauses entre les clauses étaient courtes, et le débit d'écriture plus rapide, et ce même pour les enfants de 10 ans. De même, la fréquence des mots ou leur probabilité contextuelle peuvent influer sur la temporalité de la production. Ainsi, les mots dont la probabilité contextuelle est la plus faible, qui sont généralement moins fréquents, sont davantage précédés d'une hésitation, même lorsque leur fréquence est contrôlée (Beattie & Butterworth, 1979).

L'aspect sémantique est en revanche moins important dans des tâches de production de phrases mémorisées, qui ne nécessitent donc pas de conceptualisation, puisque Ferreira (1991) ne relève par exemple pas d'effet de la plausibilité sur la latence initiale.

2.2.3.2. Aspect développemental

Des facteurs important pouvant influencer le rythme d'écriture sont la maitrise de la langue et l'automatisation des processus rédactionnels, dus notamment à l'âge des scripteurs ou locuteurs. Concernant la durée des pauses et la vitesse d'écriture, il a été relevé qu'en moyenne les pauses étaient plus longues pour les enfants de CM1 que les adultes (van Hell et al., 2008), les adultes présentant des pauses internes aux syntagmes plus courtes que les enfants, mais des pauses en début de phrase plus longues (Foulin, 1998). Par ailleurs, les enfants de 8 ans ont un débit d'écriture plus lent que ceux de 10 ans (Chanquoy et al., 1990), et les élèves de 5ème ont un débit plus lent que ceux de 3ème (Bertrand & Dumoulin, 2013).

Concernant la durée de la pause en fonction de l'emplacement syntaxique, Foulin (1998) avait comparé des productions de récits d'adultes et d'enfants de CE1 et de CE2. Les élèves de CE1 ont des pauses plus longues avant les clauses qu'avant les débuts de syntagmes. Pour les élèves de CE2, ce résultat est seulement tendanciel, mais ils font des pauses plus longues avant les débuts de syntagmes qu'avant les mots internes aux syntagmes ; les phrases, quant à elles, ne sont pas précédées d'une pause plus longue que les clauses. De même, les pauses avant les clauses étaient supérieures à celles à l'intérieur des clauses pour les élèves de CM1 de l'étude de van Hell et al. (2008). Les enfants ne marquent donc pas les unités de planification les plus élevées (paragraphes, phrases) par des pauses plus longues, ce qui peut être dû à un faible nombre de données à ces emplacements, en revanche la clause semble être une unité de planification dès le CE1. Cela ressortait également à l'oral, où des enfants de 7 ans devant imaginer la suite d'une histoire faisaient leurs pauses (supérieures à 250 ms) dans 66% des cas en frontière de clause, ces frontières rassemblant 75% du temps total de pauses de la production (Hawkins, 1971). Par ailleurs, les enfants de 8 et 10 ans de l'étude de Chanquoy et al. (1990) ne semblaient pas sensibles à la complexité syntaxique de la phrase, c'est-à-dire au fait de produire une phrase de deux clauses ou deux phrases d'une clause, ce qui pousse également à penser que la planification se fait plutôt clause par clause qu'au niveau de la phrase.

Ces différentes études suggèrent donc que l'empan de planification n'est pas le même pour les enfants, et est inférieur à l'unité « phrase ». En outre, les stratégies d'écriture diffèrent certainement, puisque la longueur des pauses diminue globalement, mais pas de la même manière suivant l'emplacement syntaxique : cela peut notamment s'expliquer par une automatisation de certains processus et par la complexification des traitements inhérente au développement de l'expertise rédactionnelle (Torrance, 1996).

3. Vers de nouvelles questions de recherche

3.1. Questions émergeant des études présentées

Les études présentées, malgré la diversité de leurs contextes de production, font donc ressortir quelques résultats récurrents. Elles présentent cependant des limites ou font émerger de nouvelles interrogations.

Tout d'abord, les études qui se concentrent sur la production de phrases suggèrent que l'empan de planification dépasse le simple mot. L'avantage de ce type de protocoles expérimentaux est qu'ils sont fortement contrôlés, et qu'ils permettent donc de tirer des conclusions sur l'effet observé. L'inconvénient est qu'ils reflètent peu ce qu'il se passe dans la production d'un texte, durant laquelle le locuteur est amené à mettre en place tous les niveaux de traitement (macro- et micro-planification, formulation, contrôle), certains traitements d'une phrase pouvant être anticipés durant la production de la phrase précédente. Il conviendrait donc davantage de mener des études au sein d'un texte ou d'un discours, mais la question de l'empan de la production a été moins exploré dans ce type d'études. Nous pouvons ici prendre appui sur les études de Fayol, Lété et Maggio (Fayol & Lété, 2012; Maggio, 2013) : ces derniers avaient constaté que les études portant sur le texte s'intéressaient principalement aux composantes de haut niveau ; après avoir mené des études sur le mot isolé (identification des effets de fréquences, de consistance, etc.), ils ont mené des études sur le mot au sein de syntagmes plus ou moins complexes, et au sein de textes. Ils ont ainsi pu voir dans quelle mesure un mot n était traité immédiatement (pendant l'écriture du mot n), mais était également anticipé (pendant l'écriture du mot n-1) ou était traité de façon retardée (pendant l'écriture du mot n+1). Concernant la micro-planification et la formulation des clauses et des phrases, après les nombreuses études portant sur la phrase isolée, il faudrait donc voir dans quelle mesure la phrase produite au sein d'un texte est anticipée ou si elle est traitée uniquement pendant sa production, voire si elle continue d'être traitée après avoir été écrite.

Au sein d'un texte, il apparait que les pauses ne sont pas réparties de manière aléatoire, mais sont plus marquées aux frontières syntaxiques ou narratives. Certains auteurs ont étudié dans quelle mesure la longueur de la pause variait en fonction du type syntaxique de clause. Les résultats, malgré quelques différences, suggèrent dans l'ensemble que la production d'une clause est plus rapide si elle est davantage dépendante (d'un point de vue syntaxique) de la clause précédente. Ainsi, il semblerait que l'empan de planification soit supérieur à la clause. Une étude plus précise serait nécessaire pour déterminer à quel niveau syntaxique se situe cet empan (clause, « phrase syntaxique », « phrase graphique »). En outre, peu d'informations sont connues concernant cette unité de production lors du développement tardif du langage (entre 10 et 15 ans), qui constitue un stade où les conventions graphiques liées à la segmentation ne sont pas tout à fait maitrisées, et où l'empaquetage syntaxique se développe.

Enfin, même si certaines hypothèses émergent quant à l'unité de production, on ne sait pas quand sont planifiées et formulées les différentes informations contenues dans les clauses, ni pourquoi ces processus sont plus ou moins longs ; seule une étude (Bertrand & Dumoulin, 2013) s'intéresse à la variation des données chronométriques en fonction du contenu sémantique des clauses. Pour répondre à cette question, il serait nécessaire de recourir à une description précise du contenu linguistique des unités et à leur mise en relation avec les données chronométriques, en utilisant les outils statistiques adéquats. Cette description du contenu devrait prendre en compte la phrase elle-même (e.g. longueur, fréquence, contenu sémantique), mais également le contenu en lien avec ce qui précède (e.g. dépendance syntaxique, accessibilité référentielle, statut informationnel), le scripteur étant à la fois influencé et contraint dans ses choix par le texte déjà-écrit (McCutchen, 2011; Plane, 2004; Plane, Alamargot et al., 2010).

Ces études ouvrent donc de nombreuses questions, tout en fournissant des résultats permettant de poser des hypothèses quant aux unités de production et à la gestion temporelle des processus rédactionnels.

3.2. Un défi méthodologique : l'analyse des données chronométriques

Dans le cadre de notre recherche, nous avons choisi d'utiliser l'enregistrement des données chronométriques, l'ordre des opérations de correction étant également enregistrées, le tout grâce au logiciel Eye & Pen. Cependant, la description des modèles ainsi que la présentation des études menées avec ce type de protocole soulignent la difficulté de mettre en relation les données chronométriques avec les différents processus observés.

En effet, pour l'écrit, les pauses ont été définies par Matsuhashi (1982, p. 270) comme des « moments d'inactivité scripturale pendant l'écriture » qui « reflètent le temps pris par les scripteurs pour engager une planification cognitive et prendre des décisions comportementales ». Cette fonction des pauses est celle ayant été principalement retenue, que ce soit à l'oral ou à l'écrit. Elle a donné lieu à plusieurs postulats, qui sont à l'origine des études évoquées précédemment et de leur interprétation, et qui selon Alamargot et al. (2007, p. 14) sont : « 1) la pause varie en fonction de la complexité des processus engagés ; 2) la position de la pause par rapport à la structure hiérarchique du texte indique la nature de ce processus ; 3) les processus qui se déroulent pendant une pause concernent la partie du texte qui sera écrite immédiatement après ; 4) lorsque les processus trop couteux ne peuvent pas être engagés en parallèle de l'exécution graphomotrice, il est nécessaire de faire une pause » [30] .

[30] notre traduction

Le problème est que, notamment à l'écrit, les pauses ne correspondent pas forcément à des moments de planification : elles peuvent également servir à la révision ou à la formulation (Foulin, 1995; Schilperoord, 2002), et Olive et al. (2009) ont montré qu'aucun des trois processus rédactionnels n'était typique des pauses, les trois intervenant à part égale dans ces moments d'arrêt scriptural. C'est ainsi que Fortier et Préfontaine (1994) ont utilisé les pauses pour étudier le mécanisme de la relecture, en effectuant un enregistrement vidéo et en classant les pauses liées à la relecture suivant leur fonction (révision de l'organisation conceptuelle ou séquentielle du texte, relecture sans modification, ou consultation du plan ou des consignes). Maggio et al. (2012) ont montré que les pauses entre deux mots n'étaient pas affectées par le mot qui allait être écrit, mais qu'en revanche elles duraient plus longtemps si le mot précédent était plus long : dans ce cas, la pause ne peut pas être seulement le corrélat d'un processus de planification, car il serait alors difficile d'expliquer pourquoi les caractéristiques d'un mot affectent la pause qui suit son écriture. On pourrait ainsi envisager que la pause longue entre deux phrases serve à intégrer en MLT la phrase qui vient d'être écrite (Daiute, 1984). Enfin, les pauses ne sont pas toujours le reflet de processus rédactionnels. Certaines pauses, relativement courtes, peuvent avoir une cause physique, comme une fatigue musculaire, ou la nécessité de passer à la ligne suivante (dans le cas de l'écriture manuscrite). D'autres facteurs peuvent conduire à des pauses plus longues : le scripteur peut être perturbé par du bruit et arrêter sa rédaction, regarder l'heure, penser à autre chose, etc. À l'oral, d'autres fonctions peuvent être associées à la pause, comme la nécessité de reprendre son souffle (Goldman et al., 2010), d'insister sur un point (Degand & Simon, 2009) ou de marquer une frontière perceptible par l'auditeur (Oliveira, 2002b).

En outre, le problème de savoir quelles « pauses » sont étudiées demeure. Il est ainsi possible de choisir un seuil, et d'étudier tous les moments où la transcription est arrêtée pendant une durée supérieure à ce seuil, comme l'a fait Piolat (1983). Cependant, le choix du seuil n'est pas toujours justifié, et un seuil trop élevé (par exemple 1 seconde) peut amener à ne pas prendre en compte de nombreuses pauses qui seraient pourtant liées à l'activité cognitive (Chenu, Pellegrino, Jisa & Fayol, 2014). En outre, il pourrait être davantage pertinent d'établir un seuil individuel pour chaque scripteur, dépendant de ses compétences d'exécution ou de son style d'écriture, plutôt qu'un seuil général (Baaijen et al., 2012; Olive, 2011). D'autres chercheurs considèrent que les pauses sont définies comme le temps séparant deux mots (Chanquoy et al., 1996). Si cela semble plus objectif, des questions peuvent être soulevées : dans le cas où deux mots sont séparés par un signe de ponctuation, doit-on additionner le temps avant et après le signe, la répartition temporelle de la pause avant et après le signe pouvant varier en fonction du signe de ponctuation et du scripteur (Foulin, Chanquoy & Fayol, 1989) ? Par ordinateur, la pause entre deux mots est-elle l'addition de la pause avant la barre « espace » et la pause avant le mot (Baaijen et al., 2012) ? Cette définition a en outre ses limites, puisqu'il est aussi possible d'étudier les pauses intra-mots (Maggio et al., 2012). Une nouvelle piste de recherche pour étudier les pauses, qui prendrait notamment en compte la variation interindividuelle, serait d'étudier les points de ruptures de la dynamique d'écriture (Chenu et al., 2014).

Enfin, le dernier problème serait de limiter l'analyse des processus cognitifs à l'étude des pauses. En effet, l'étude d'Olive et al. (2009) montre que, en même temps que la transcription graphique, d'autres processus rédactionnels sont engagés : pendant qu'ils écrivaient, il arrivait aux scripteurs de réfléchir à la formulation ou à la planification ; par moments, ils effectuaient même des révisions en même temps que la transcription. Cela inciterait donc à considérer le débit comme pouvant être également une source d'information sur les processus cognitifs engagés lors de la production écrite. Van Galen (1991) soulignait déjà que la transcription graphique n'était pas simplement un mécanisme biologique, mais que les mouvements d'écriture pouvaient être le reflet de mécanismes cognitifs, notamment liés aux niveaux lexical et sémantique. Maggio et al. (2012) ont ainsi vu que des variables associées au mot (position du mot dans le texte, fréquence du mot, fréquence des syllabes, consistance orthographique) affectaient les pauses intra-mots et le débit d'écriture.

Les études que nous allons mener sur les processus rédactionnels en nous appuyant sur la temporalité de l'écriture nous conduiront à nous interroger sans cesse sur les indices utilisés, et l'interprétation que nous en faisons. Des terminologies différentes existent pour désigner ce qui laisse une marque et doit être interprété : « trace », « empreinte », « indice », « signe » (Doquet-Lacoste, 2004; Jeanneret, 2010; Krämer, 2010; Serres, 2002). Les définitions sont précises, et les termes ne recouvrent pas exactement la même chose suivant les auteurs. Dans le cadre de notre étude, nous utiliserons le terme de « trace » pour désigner « un objet inscrit dans une matérialité que nous percevons dans notre environnement extérieur et dotons d'un potentiel de sens particulier » (Jeanneret, 2010, p. 61). L'auteur de cette définition ajoute que ce sens est « la capacité dans le présent de faire référence à un passé absent mais postulé » ; cette restriction correspond à ce que nous appellerons « empreinte ». Nous ajouterons que la trace peut servir à renvoyer à « des choses qui arrivent en même temps sans être pour autant visibles en même temps » (Krämer, 2010, p. 14) ; il sera alors question d'« indice ».

Il faudra cependant toujours préciser « de quoi » un élément est la trace. Dans nos textes, nous avons peu d'empreintes. Nous pouvons mentionner les modifications apportées au texte : celles-ci sont l'empreinte d'un retour sur le texte, et d'une modification de ce texte. Il faudra s'interroger sur la manière dont ce type de traces peut également être l'indice d'autres processus rédactionnels. Les données chronométriques, quant à elles, ne constitueront que des indices : indice d'une planification ou formulation plus ou moins couteuse, indice d'une relecture. Pour interpréter au mieux ces données chronométriques, il sera nécessaire de prendre en compte de manière fine le contenu linguistique entourant les périodes analysées. Par exemple, pour mieux appréhender les processus de micro-planification et de formulation, il sera possible de mettre en relation les données chronométriques (pauses initiales, moyenne des pauses intra-mots au sein d'une clause ou d'une phrase, moyenne des pauses inter-mots, moyenne de la vitesse d'écriture) avec le contenu linguistique (e.g. continuité référentielle, type de prédicat, fréquence lexicale) de ce qui est produit ou de ce qui va être produit.

Concernant les révisions, il est possible d'analyser d'une part le produit final (e.g. le niveau linguistique affecté, le type de correction) pour comprendre l'intention du scripteur, et d'autre part d'effectuer une analyse temporelle (e.g. localisation dans le texte de l'endroit où la révision est initiée et du segment affecté, durée de la session de révision, durée des pauses durant la révision). À une autre échelle, nous attendons que les observations quant au cout et à la localisation de ces divers processus soient à leur tour des indices concernant les unités de production au cours de l'écriture d'un texte.

Toutes ces considérations inviteront donc à la prudence lors des interprétations de nos résultats, ou permettront de proposer des interprétations diverses à ce que nous observerons. Afin d'éviter un maximum de biais d'interprétation dans ce qui sera étudié, il sera nécessaire de recourir à des analyses rigoureuses.

3.3. Présentation de notre recherche

La présentation des modèles, des outils à notre disposition ainsi que des études portant sur le lien entre durée des pauses et contenu linguistique a fait émerger de nombreuses questions.

La problématique de notre recherche, qui sera à l'origine des différentes études que nous présenterons, est la suivante : quelle est la taille des unités de micro-planification et de formulation et comment la production de ces unités est-elle gérée d'un point de vue temporel ? Ainsi, nous examinerons dans quelle mesure la temporalité de l'écriture nous renseigne sur les unités de production lors de l'écriture d'un texte.

Cette problématique se déclinera en plusieurs questions :

- Un pattern développemental concernant les unités de production apparait-il ?
- Les adultes anticipent-ils la planification et la formulation des clauses dépendantes ?
- Quelles caractéristiques linguistiques d'une unité rendent sa micro-planification ou sa formulation plus couteuse ?
- Comment les révisions sont-elles gérées au sein des unités de production ?

Pour répondre à ces questions, nous utiliserons deux corpus : un corpus développemental (avec des productions d'élèves de CM2, 5ème et 3ème) et un corpus de jeunes adultes. Les données chronométriques prises en compte seront les pauses entre les mots, différenciées suivant leur emplacement syntaxique, les pauses intra-mots et le temps d'exécution. Afin d'étudier la gestion temporelle précise de la micro-planification et de la formulation, nous fournirons une description fine des caractéristiques sémantiques, syntaxiques et lexicales de chaque clause et chaque phrase. Celles-ci seront ensuite mises en relation avec les données chronométriques de l'unité en cours d'écriture et de l'unité précédente : grâce à des outils statistiques adaptés, il sera possible de voir quels facteurs linguistiques expliquent la variance de

ces données chronométriques, permettant de préciser à quels moments se déroulent la micro-planification et la formulation. Enfin, l'étude des révisions apportera des indices supplémentaires quant aux unités de production et au rôle de la relecture lors de la production écrite.

Le chapitre suivant présentera les aspects généraux de la méthodologie : constitution des corpus, codage général, analyses utilisées. Il sera suivi de quatre chapitres (chapitres 5 à 8) présentant les résultats de nos études. Le chapitre 5 portera sur le corpus développemental ; il interrogera la question des unités de production chez les enfants et adolescents, et cherchera à mettre en lien les unités de productions avec la maitrise de la segmentation du texte. Les chapitres 6 et 7 seront basés sur les données chronométriques du corpus adulte. Nous présenterons tout d'abord la gestion chronométrique des textes, en soulignant les différences interindividuelles présentes, et nous verrons dans quelle mesure le lien entre données chronométriques et empaquetage syntaxique permet de faire ressortir les unités de production (chapitre 6). La gestion temporelle de la micro-planification et de la formulation des clauses et des phrases, liée aux différences de cout cognitif en fonction du contenu linguistique, sera ensuite précisée (chapitre 7). Enfin, le chapitre 8 s'appuiera sur l'analyse des révisions du corpus adulte pour vérifier certaines hypothèses concernant les unités de production ; une étude de cas permettra en outre d'interroger à nouveau la manière d'utiliser les indices à notre disposition, et ouvrira la question du rôle de la relecture dans les processus de planification et de formulation.

Des hypothèses précises précèderont chaque partie de cette étude, mais nous pouvons évoquer quelques hypothèses générales :

- Concernant l'aspect développemental, il est attendu que la notion de « phrase » soit de plus en plus pertinente du point de vue cognitif comme unité de production.
    - Les élèves les plus jeunes planifieraient davantage « clause par clause » qu'à un niveau supra-clausal.
    - L'appropriation de l'unité « phrase » serait liée à la manière dont elle est enseignée, notamment par le biais des marques typographiques de segmentation.
- Chez les adultes, il est attendu qu'une partie des phrases soient planifiée à un niveau supra-clausal.
    - Plus l'emplacement syntaxique est élevé dans la hiérarchie, plus la pause devrait être longue.
    - Une partie du contenu sémantique des phrases serait anticipée dès l'écriture de la fin de la phrase précédente.
    - Les clauses subordonnées seraient plus particulièrement planifiées lors de l'écriture de la clause matrice.
    - Le contenu linguistique influerait sur la dynamique d'écriture : par exemple, des éléments plus fréquents ou plus faciles d'accès seraient plus rapides à produire ou à

planifier, alors que des structures plus complexes ou plus longues seraient plus longues à produire ou à planifier.

- La temporalité de l'écriture ne saurait être totalement appréhendée en n'étudiant que le contenu linguistique des unités en cours de production.
    - L'étude des révisions dans leur aspect temporel fournirait de précieux indices quant aux processus de micro-planification et de formulation.
    - Plus particulièrement, les retours sur le texte n'ayant pas vocation à le corriger indiqueraient l'importance de la relecture du texte déjà écrit au fur et à mesure de la production de celui-ci.

Méthodologie générale

Chapitre 4 – Méthodologie générale

Nos analyses porteront sur deux corpus de données : le corpus « Reflex » et le corpus « Dimitri » ; nous détaillerons la manière dont ces données ont été recueillies. Nous présenterons ensuite le codage appliqué de manière similaire aux deux corpus : d'un point de vue linguistique, le texte a été segmenté en différentes unités, et les liens syntaxiques entre les clauses ont été indiqués ; d'un point de vue des données chronométriques, des variables ont été retenues pour chaque unité, et nous indiquerons certains choix effectués quant au traitement de ces données. Enfin, nous expliquerons les différents types d'analyses qui seront utilisés pour nos études.

1. Recueil et traitement des données

Nos données sont issues de deux corpus, collectés dans le cadre de projets menés au sein du laboratoire Dynamique du Langage, en collaboration avec des équipes de chercheurs en Psychologie cognitive (LAPSCO, EMC, CerCa, CHArt) : le projet ReFlex[31] et le projet DyTO[32]. Le premier corpus nous permettra de répondre à nos questions concernant l'aspect développemental des unités de production. Ce corpus, qui avait été collecté en 2005-2006, est en effet constitué de production d'enfants et adolescents de 10 à 15 ans, qui permettent donc d'explorer la question du développement des unités de production. Cependant, des limites sont apparues quant à l'utilisation de ce corpus ; celles-ci seront décrites à la fin du chapitre 5. Par conséquent, afin de répondre de manière plus précise à nos questions concernant la gestion temporelle des unités de production, nous avons eu recours à un autre corpus, collecté en 2012, rassemblant des textes produits par des jeunes adultes. Nous allons à présent donner les caractéristiques de ces deux corpus : participants, tâches demandées, matériel utilisé.

1.1. Recueil des données du corpus « ReFlex »

Pour la première partie de notre travail, nous avons utilisé un corpus déjà existant, collecté en 2005-2006 dans le cadre de l'ANR ReFlex. Le détail du recueil de données est disponible dans la thèse de Mazur-Palandre (2009). Nous ne donnerons ici que quelques informations nécessaires à la compréhension du type de données à notre disposition.

1.1.1. Les participants

Les données ont été recueillies auprès de 132 élèves de CM2 (N=44), 5ème (N=45) et 3ème (N=45). Tous les participants étaient des francophones natifs monolingues, ne présentaient pas de retard ou d'avance scolaire supérieurs à un an, et n'avaient pas de troubles du langage comme la dyslexie. Pour la spécificité du projet ReFlex, les données ont été recueillies dans des établissements situés en ZEP (Zones d'Éducation Prioritaire). De plus, à partir d'un questionnaire mis en lien avec des documents de l'INSEE, seuls des enfants appartenant à la classe sociale défavorisée ont été sélectionnés.

Dans notre étude, nous avons conservé les productions de 40 élèves par niveau scolaire, en nous assurant de conserver la parité filles / garçons pour chaque groupe.

---

[31] Projet n° NT 05-2_41686 « Renforcer la flexibilité lexicale et syntaxique dans l'expression orale et écrite des adolescents » ; porteur du projet, Harriet Jisa (Laboratoire DDL, UMR 5596, Université de Lyon) ; coordinateur, Florence Chenu (Laboratoire DDL, UMR 5596, Université de Lyon) ; collaborateur principal, Michel Fayol (Laboratoire LAPSCO, UMR 6024, Université Blaise Pascal)

[32] Projet ANR-11-BSH2-0002 « Dynamique des Traitements Orthographiques » ; porteur du projet, Denis Alamargot (Laboratoire CHArt, UPEC) ; coordinateur, Bernard Lété (Laboratoire EMC, Université de Lyon)

1.1.2. Tâches demandées

Il était demandé aux participants de produire quatre textes, puisqu'il y avait deux modalités (oral / écrit) et deux types de texte (narratif / expositif). Pour les textes écrits, les élèves produisaient une première version « brouillon », puis une version « propre ». Tous les textes abordaient la thématique des « problèmes entre les gens ». Le protocole était inspiré du projet Spencer, une étude psycholinguistique internationale. Une vidéo avait été créée pour ce projet : elle présentait une série de petites scènes portant sur des problèmes liés à l'école, comme la tricherie, l'exclusion, le vol, les disputes ; il s'agissait d'une vidéo muette, afin de pouvoir être utilisée dans le cadre de ce projet international.

Chaque enfant devait donc produire des textes de deux types : d'une part « un exposé sur les problèmes entre les gens », et d'autre part une « histoire personnelle » racontant ce type de problèmes. Les consignes permettant d'amorcer la production sont présentées dans la Figure 7.

Texte narratif :
« Tout à l'heure, tu as vu différentes sortes de problèmes. J'aimerais que tu me racontes un problème (ou une dispute) que tu as eu avec quelqu'un. Ne me rapporte pas ce que tu as vu dans la vidéo mais raconte une histoire personnelle qui t'est arrivée. Tu peux prendre ton temps. »
Texte expositif :
« Tout à l'heure, tu as vu différentes sortes de problèmes. J'aimerais que tu parles sur ce sujet. Discute et présente tes idées comme si tu étais debout devant une classe. Dis-moi ce que tu connais concernant ce genre de problèmes. Ne raconte pas une histoire mais discute du sujet (ou du thème). Tu peux prendre du temps pour réfléchir et quand tu es prêt, commence à parler. »

Figure 7. Consignes données pour la production des textes - corpus « ReFlex »

Les données ont été collectées en deux sessions pour chaque élève : lors de la première session, les participants ont regardé la vidéo puis ont rédigé les deux textes, qui différaient simplement par la modalité ; la seconde session a eu lieu une semaine après, et ils devaient alors rédiger l'autre type de texte, de nouveau dans les deux modalités. Lors de chaque session, les deux productions étaient séparées par un questionnaire auquel le participant répondait oralement. Afin de contre-balancer l'ordre de production, pour le type de texte et la modalité, quatre groupes ont été établis. Le Tableau 2 indique l'organisation des sessions. Ainsi, les participants des groupes A et B ont d'abord composé le texte narratif, puis la semaine suivante le texte expositif, contrairement aux participants des groupes C et D ; en outre, les participants des groupes A et C ont toujours débuté leurs sessions par l'oral puis ont produit le texte écrit, à l'inverse des participants des groupes B et D.

	Semaine 1	**Semaine 2**
A	Narratif oral – Narratif écrit	Expositif oral – Expositif écrit
B	Narratif écrit – Narratif oral	Expositif écrit – Expositif oral
C	Expositif oral – Expositif écrit	Narratif oral – Narratif écrit
D	Expositif écrit – Expositif oral	Narratif écrit – Narratif oral

Tableau 2. Ordre de passage lors du recueil des données « ReFlex »

Pour notre étude, nous avons uniquement conservé les productions écrites, soit 240 textes (2 types de textes x 40 participants x 3 niveaux). En outre, comme notre travail s'intéresse davantage aux processus rédactionnels qu'au produit final, c'est la version « brouillon » qui a été utilisée, et non la version « propre ».

1.1.3. Matériel utilisé

Les productions orales, qui ne seront pas étudiées dans le cadre de cette thèse, ont été enregistrées grâce à des enregistreurs Edirol R-1.

Les productions écrites ont été collectées grâce à une tablette graphique Wacom Intuos 3 A4 oversize, reliée à un ordinateur portable. Les participants écrivaient sur une feuille de papier A4 similaire à celles utilisées en classe, c'est-à-dire avec une marge et des interlignes ; ils utilisaient un stylet possédant une vraie mine d'encre. Ainsi, cela se rapprochait des conditions de rédaction habituelles, avec une feuille de papier et un stylo classique. Les données, transmises à l'ordinateur, étaient ensuite enregistrées grâce au logiciel Eye & Pen (Chesnet & Alamargot, 2005), qui collectait les données spatiales et chronométriques relatives à la production.

1.2. Recueil des données du corpus « Dimitri »

La deuxième partie de notre travail sera basée sur un corpus que nous avons collecté en 2012 auprès de jeunes adultes. Seule la tâche pendant laquelle les scripteurs devaient raconter l'histoire d'un personnage nommé Dimitri sera utilisée dans notre étude ; c'est pourquoi nous nommons ce corpus « Dimitri ».

1.2.1. Les participants

Les participants étaient tous de jeunes adultes, généralement étudiants ou doctorants, de langue maternelle française. Les textes ont été écrits par 31 personnes, dont 8 garçons, qui avaient en moyenne 22 ans et 10 mois. Ce recueil de données s'est déroulé entre mars et juillet 2012. Un texte s'est particulièrement distingué des autres : de nombreuses corrections ont été apportées, la participante écrivant même dans la marge et à l'extrémité basse de la feuille. Compte tenu du type d'analyses que nous voulions mener, nous avons écarté ce texte du corpus, qui comptera donc 30 productions. En revanche, ce texte constituera une étude de cas, présentée à la fin de cette thèse.

1.2.2. Tâches demandées

Il était demandé aux participants d'écrire deux récits : « Chaperon » et « Dimitri », dont nous allons à présent expliciter plus précisément les consignes. La moitié des participants a commencé par la tâche « Chaperon » (ordre A), et l'autre moitié par la tâche « Dimitri » (ordre B).

Lors de la tâche « Chaperon », cinq images relatant les principales péripéties de l'histoire du « Petit Chaperon Rouge » étaient présentées au participant. Ce dernier devait alors raconter le conte dans son ensemble, à partir des images et de ses souvenirs. La fin de l'histoire n'étant pas décrite par les images, il était nécessaire de faire appel à ses souvenirs ou d'imaginer le dénouement final. La Figure 8 expose les cinq images présentées au participant.

Figure 8. Stimulus utilisé pour la tâche « Le Petit Chaperon Rouge »

Lors de la tâche « Dimitri », le participant devait écrire la suite d'une histoire qui soit en cohérence avec le début proposé, dont le titre était *L'évasion*. L'énoncé servant d'amorce à la suite du récit est présenté dans la Figure 9.

L'évasion
Dimitri regardait le ciel étoilé à travers les barreaux. Il devinait les ombres des gardiens dans le mirador. Il entendit la porte de sécurité s'ouvrir dans le couloir. Le gardien ouvrit la porte de sa cellule. Il passa alors à l'action.

Figure 9. Texte d'amorce proposé pour la tâche « Dimitri »

Les deux récits devaient être composés d'une suite de péripéties allant vers un dénouement final. Il était demandé aux participants que le texte composé ne soit pas trop complexe, donnant comme indication qu'il devait être compréhensible pour un enfant de CM2.

Dans le cadre de cette thèse, seuls les textes « Dimitri » ont été utilisés. En effet, le but de ce travail est de voir de manière très précise comment les processus de micro-planification et de formulation sont gérés par le scripteur. Or, ceux-ci diffèrent très certainement dans le cas d'une histoire à inventer (du type de la tâche « Dimitri ») et dans le cas d'une histoire à se remémorer (du type de la tâche

« Chaperon »). De plus, pour certaines de nos analyses, nous avons utilisé des informations linguistiques dont l'influence n'est vraisemblablement pas la même d'un type de texte à l'autre : par exemple le type de sujet, le sémantisme des prédicats, etc. Pour plus d'homogénéité, nous avons donc préféré nous en tenir à un seul type de texte, l'invention de la suite d'un récit fictif, représenté par les textes de la tâche « Dimitri ».

1.2.3. Matériel utilisé

Le matériel utilisé est le même que pour le recueil du corpus « ReFlex », à savoir une tablette graphique Wacom Intuos 3 A4 oversize, reliée à un ordinateur portable. Le même type de stylo était utilisé ; les participants écrivaient sur une feuille de papier de format 30x30 cm. La différence principale entre les deux collectes concerne la version du logiciel utilisé : pour ce recueil de données, il s'agit de la version 2 du logiciel Eye & Pen, qui n'était pas encore disponible lors du recueil du corpus « ReFlex ». Cependant, cette différence concerne essentiellement la manière de coder et d'extraire les données à partir du logiciel, et non le recueil en lui-même.

1.2.4. Déroulement de l'expérience

Une première phase de préparation se déroulait avant l'arrivée du participant. L'expérimentateur disposait deux feuilles 30x30 cm sur la tablette graphique, et laissait le stylet à droite de la tablette. Les supports des tâches (images et début de récit) étaient cachés afin de n'être dévoilés qu'avant chaque tâche.

Le participant était installé face à la tablette, qui était fixée sur la table. L'expérimentateur lui montrait le matériel, et lui proposait d'écrire son nom et la date sur une feuille vierge, afin de se familiariser avec la tablette et le stylet. Les participants contribuant à l'expérience de manière anonyme, cette feuille était tout de suite jetée. L'expérimentateur exposait ensuite brièvement le but de l'expérience, puis précisait au participant les consignes générales : il devrait produire deux récits, l'un racontant une histoire connue, et l'autre la suite d'une histoire qui lui serait proposée, ces récits devant faire chacun au minimum 15 lignes (une demi-page environ), et au maximum une page.

L'expérimentateur lisait ensuite la consigne de la première tâche et présentait le support. Il démarrait l'enregistrement en appuyant sur une touche de l'ordinateur. Après avoir donné le stylet au participant, il lui demandait d'inscrire une croix en haut de la feuille dès le début de sa réflexion. Cela permettait notamment de vérifier le bon fonctionnement de l'enregistrement des données. La première tâche terminée, le participant appuyait avec son stylet sur une zone de sa tablette où il était indiqué « Suite », ce qui arrêtait l'enregistrement de cette tâche. L'expérimentateur enlevait la feuille du premier récit, sous laquelle était déjà disposée une feuille vierge, et procédait de la même manière pour la deuxième tâche. La fin de la seconde production était marquée par l'appui du participant sur la zone « Fin » de sa tablette.

1.3. Extraction des données

Le logiciel Eye & Pen permet d'enregistrer les mouvements du crayon des scripteurs durant les tâches de production. Il est ainsi possible de « rejouer le film » de la production ; en outre, le type d'évènement (pause levée, pause appuyée, stylo en mouvement) et la durée de ces évènements sont enregistrés. À partir de ces enregistrements, il est tout d'abord possible de segmenter le texte afin de faire ressortir des « séquences ». Des traitements permettent ensuite, à partir de l'extraction des données, d'aligner les données chronométriques sur les séquences identifiées (e.g. pause initiale, somme des pauses internes, durée d'exécution).

1.3.1. Séquençage des mots

Afin d'utiliser les données collectées par le logiciel, la première étape a consisté à créer des séquences : c'est ce que l'on nomme le « détourage » de mots. Dans la version 1 du logiciel, ce détourage est entièrement manuel. Dans la version 2, les séquences sont créées de manière semi-automatique grâce à l'outil de détourage des mots, en fonction de paramètres que l'utilisateur ajuste à l'écriture de chaque sujet, puis les séquences sont affinées par l'expérimentateur. Une séquence correspond à un « mot graphique », c'est-à-dire à une chaine de caractères délimitée par deux blancs ; elle débute lorsque la pause qui précède le mot commence, et se termine au début de la pause précédant le mot suivant. Chaque signe de ponctuation correspond également à une séquence. Lorsque les mots sont détourés, plusieurs éléments sont indiqués par le codeur : les majuscules, les erreurs d'orthographe, les révisions effectuées, les retours à la ligne ou les débuts de paragraphe. La Figure 10 donne un exemple de détourage des mots d'un texte.

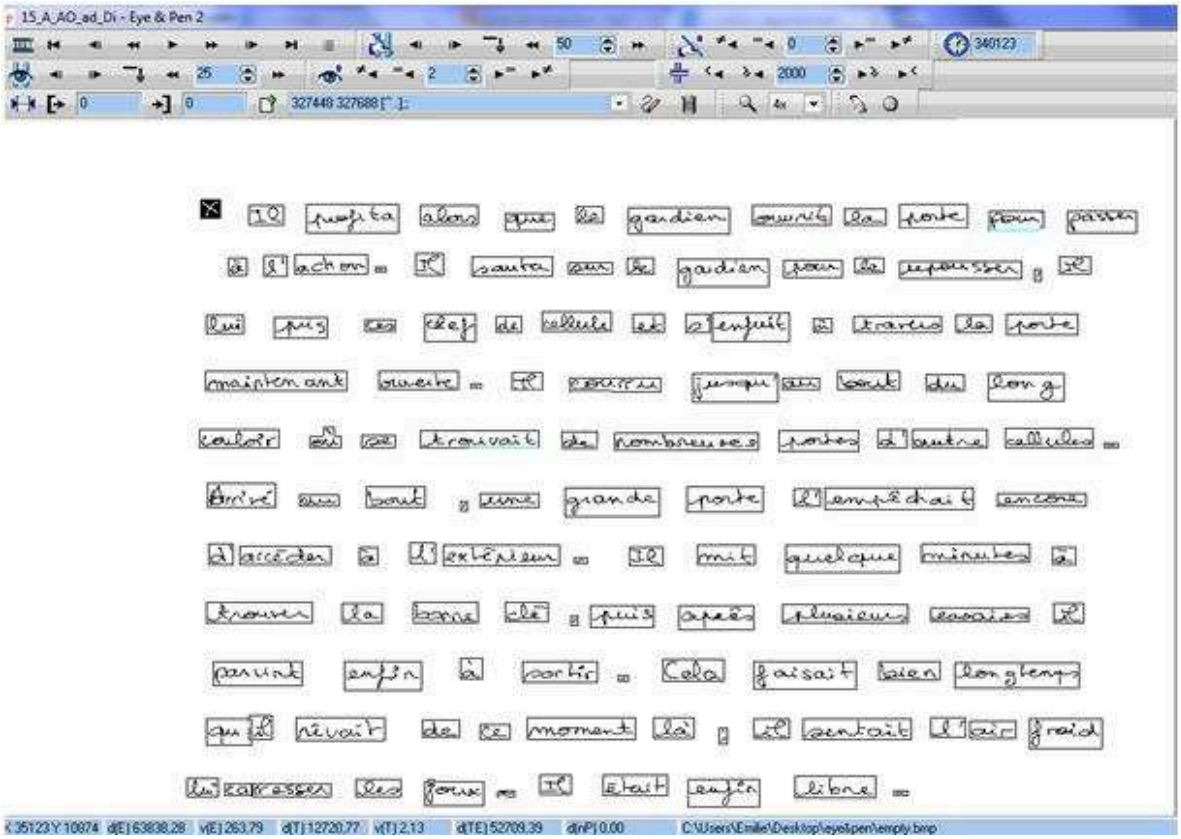


Figure 10. Exemple de détourage des séquences d'un texte avec le logiciel Eye & Pen

1.3.2. Extraction des données chronométriques

Le logiciel Eye & Pen permet d'extraire un certain nombre de données, qui sont ensuite traitées par des scripts afin d'associer à chaque séquence les données chronométriques qui nous intéressent, dont l'unité est la milliseconde (ms). Tout d'abord, il y a la « pause initiale », correspondant à la pause précédant le mot. Les « pauses intra » ont également été comptabilisées : elles correspondent à la somme des pauses effectuées au cours de l'écriture de la séquence, divisée par le nombre de caractères du mot. Enfin, le « temps d'exécution » du mot, correspondant au temps mis pour écrire le mot, après exclusion des pauses-intra, divisé par le nombre de caractères, a également été calculé. Une « pause » est relevée à partir du moment où il y a une inactivité scripturale, avec un seuil minimal de 15 ms (qui était le seuil minimal de la 1ère version de Eye & Pen).

2. Codage général

La problématique générale de notre travail est de savoir quelles sont les unités de production et comment celles-ci sont gérées temporellement. Pour cela, la méthode retenue est la mise en relation des données chronométriques et du contenu linguistique des textes. Dans un premier temps, un codage commun a donc été effectué pour répondre aux premières questions, à savoir l'empan des unités de production. D'une part, il a été codé au niveau linguistique plusieurs types de segmentation du texte, ainsi que des informations syntaxiques permettant de spécifier les liens entre les unités. D'autre part, des choix ont dû être faits concernant la manière dont les données chronométriques seraient utilisées ; chaque variable chronométrique sera en outre précisément décrite.

2.1. Données linguistiques : les grandes unités de segmentation et leurs relations

2.1.1. Les unités de segmentation du texte

Un des enjeux de cette thèse est d'identifier des « unités de productions » à l'écrit. Dans la démarche adoptée, il est nécessaire de segmenter le texte en différentes unités, pouvant potentiellement être ces unités de production. Quatre grands « niveaux » ont été retenus : la phrase graphique, l'Unité Terminale, l'Unité Autonome et la clause. Ces quatre niveaux n'ont pas toujours été inclus dans nos analyses. Les niveaux intermédiaires « Unité Terminale » et « Unité Autonome » sont notamment proches : ils suivent des critères syntaxiques en regroupant une proposition principale et les propositions subordonnées. La différence entre ces deux niveaux concerne le traitement réservé aux propositions principales présentant une ellipse du sujet : elles forment une seule « Unité Terminale » avec la proposition principale précédente, à laquelle elles sont « rattachées », mais elles forment une « Unité Autonome » distincte. Le choix entre ces deux unités diffèrera selon les études : l'Unité Terminale sera utilisée dans les études développementales, conformément aux précédentes études portant sur l'empaquetage syntaxique pour ce corpus, tandis que l'Unité Autonome, plus opératoire pour décrire en détail le contenu linguistique des phrases, sera utilisée dans les études portant sur le corpus « Dimitri ».

2.1.1.1. La phrase graphique (« P »)

Selon les définitions des grammaires traditionnelles, une phrase est « une unité de sens accompagnée […], à l'écrit, par les signes typographiques que sont, en français, la majuscule et le point » (Dubois et al., 2007, p. 362). Nous avons vu dans le premier chapitre la difficulté de donner une définition linguistique consensuelle de la « phrase », conduisant à des difficultés d'enseignement, et donc d'apprentissage. Même si cette unité peut être critiquée de par la multitude de critères pouvant être invoqués pour la définir, nous avons choisi de la retenir, en prenant en compte le critère graphique. Dans

nos corpus, il sera donc considéré qu'une unité commençant par une majuscule et/ou[33] se terminant par un point est une phrase. Ainsi, dans cette étude, le terme de « phrase » renverra systématiquement à la phrase graphique.

2.1.1.2. L'Unité Terminale (« T »)

Devant la difficulté de définir une « phrase », il pourrait paraitre plus pertinent de définir une unité à partir de critères syntaxiques, cette unité pouvant en outre être utilisable à l'oral. C'est ce qu'a entrepris Hunt, en 1970, lorsqu'il a défini la minimal terminable unit comme étant l'unité minimale qui pourrait être ponctuée comme une phrase tout en étant grammaticalement correcte.

L'unité Terminale (unité T) est donc définie comme une unité qui contient une proposition matrice ainsi que toutes les clauses dépendantes : clauses subordonnées (1) et clauses juxtaposées ou coordonnées avec ellipse (2).

(1) *Il sortit de sa cachette / dès qu'il n'entendit plus de bruit*[34]*. ///* [S17][35]

(2) *Il parcourut à peine six mètres / et trouva l'emplacement idéal. ///* [S28]

Par ailleurs, les travaux menés notamment sur l'oral ont mis en avant l'existence de structures qui, bien que ne répondant pas à l'idée classique de « clause indépendante », pouvaient être considérées comme des unités Autonomes. Foster, Tonkyn et Wigglesworth (2000) ont ainsi proposé l'« AS-unit », qui inclut, en plus des clauses indépendantes, des unités sous-clausales. Ces dernières sont souvent composées d'éléments ellipsés, d'où leur nature nominale, et se caractérisent à l'oral par une pause prosodique. Ces travaux nous ont menés à considérer comme unité T toute nouvelle phrase, même si elle ne contient pas de verbe (3), la présence d'une ponctuation forte montrant la volonté de rendre l'unité indépendante.

(3) *Tout s'était passé comme prévu. ///* Aucune surprise. *///* [S21]

Enfin, un dernier point problématique pour définir les unités T concernait le discours rapporté direct introduit ou suivi par un verbe de parole. En effet, un tel verbe, comme « dire », nécessite un complément, qui peut apparaitre sous forme de discours direct. Il s'agit alors d'une clause dépendante, que l'on peut donc considérer comme faisant partie de la même unité T terminale que la clause principale. Cependant, lorsque le discours direct est composé de plusieurs phrases, il est difficile de

[33] Certains scripteurs, notamment les plus jeunes, oubliant parfois de mettre une majuscule en début de phrase ou de la terminer par un point, seule la présence d'un des deux critères était nécessaire pour délimiter une phrase graphique.

[34] Le signe "/" indique une séparation entre deux clauses ; le signe "///" une séparation entre deux unités T.

[35] Les exemples sont principalement issus de notre corpus d'étudiants à l'université. Dans ce cas, seul le numéro du sujet, servant à l'identifier, est indiqué entre crochets.

regrouper l'ensemble dans une même unité T. La solution la plus adaptée nous semble être celle de O'Hare (1973), rattachant la première phrase à l'unité T contenant le verbe introducteur, et décomptant les unités T suivantes de manière indépendante : ainsi, dans l'exemple (4), nous comptons trois unités T (délimitées par « /// »).

(4) *Les seules réponses /[qu'il avait]/ c'était celles des autres détenus, /qui criaient* / : « Vas-*y fonce petit, / sors vite d'ici. /// On a confiance en toi, /// tu vas y arriver.* » [S26][36]

2.1.1.3. L'Unité Autonome (« A »)

L'unité T a permis d'isoler des unités qui seraient syntaxiquement correctes si elles étaient ponctuées comme des phrases. Cependant, d'un point de vue sémantique, ces unités peuvent contenir plusieurs prédicats qui ne sont pas dans une relation de subordination. Afin de segmenter le texte en unités sémantiques principales centrées autour d'un prédicat unifié, nous avons introduit l'unité Autonome (unité A) : il s'agit ici d'une proposition matrice avec toutes ses clauses subordonnées. La différence avec l'unité T est que les propositions principales dont le sujet est une ellipse constituent une unité A indépendante (5).

(5) Il parcourut à peine six mètres // *et trouva l'emplacement idéal*. // [S28][37]

2.1.1.4. La clause (« C »)

Comme indiqué au chapitre 1 de cette thèse, nous avons choisi de définir la notion de « clause » à partir de ce qu'ont proposé de Berman et Slobin (1994, p. 7) : « toute unité qui contient un prédicat unifié ». Cette notion de « prédicat » n'est pas aisée à définir, c'est pourquoi il est nécessaire d'expliciter la manière dont nous avons segmenté le texte en clauses.

La plupart du temps, le prédicat est verbal. Afin de définir les limites de ce qu'est un prédicat verbal, nous avons suivi en grande partie la proposition de Gayraud (2001). Le détail de la segmentation en clauses de ce corpus est disponible en Annexe 1. Les choix de segmentation en clauses sont résumés dans le Tableau 3.

[36] Les clauses enchâssées sont indiquées entre crochets.
[37] Le signe "//" indique une séparation entre deux unités A.

Verbes	Type de sujet	Critère	Nombre de clauses	Exemples
1 verbe fléchi dans l'unité A			1 clause	*il connaissait ces couloirs par cœur* [S23]
1 verbe fléchi + 1 verbe non fléchi	Sujets coréférentiels	Verbe à l'infinitif sature la valence verbale	1 clause	Sous le choc, le gardien ne pouvait pas se relever [S06]
		Autres	2 clauses	Dimitri sortit / pour aller diner. [S10]
	Sujets non coréférentiels	Structure causative	1 clause	et les laissa tomber par terre [S03]
		Autres	2 clauses	il demanda alors à son ami le gardien / de faire diversion [S05]
2 verbes fléchis dans l'unité A		« il y a » marquant la durée	1 clause	Dimitri repensa à son arrivée il y a de ça *presqu'un an* [S28]
		Groupe nominal clivé avec « c'est… qui /que »	1 clause	*C'est à ce moment*-là que Dimitri élança sa jambe vers la tête du gardien [S01]

Tableau 3. Segmentation en clauses – Résumé des critères

2.1.1.5. Exemple de segmentation

Afin d'illustrer ce codage de segmentation du texte, nous présentons l'exemple (6), qui est un texte segmenté en phrases, unités T, unités A et clauses [38].

(6) ~ Le couteau [dérobé le midi même à la cantine] était tenu fermement par sa main droite ; /// *d'un pas rapide, il se plaça devant le gardien, // lui mit la lame sous la* gorge // et lui intima / d*'être silencieux. ~ Il ne s'agissait pas / d'ameuter le service de sécurité au complet dès le commencement de l'action.* ~ Il demanda au gardien / *d'avancer plus en avant dans la cellule / et de s'assoir sur le lit.* ~ Il murmurait tous *ces mots / qu'il s'ét*ait déjà répétés cent fois. ~ *Oui, il avait planifié jusqu'aux mots / qu'il dirait au gardien, /// il ne fallait / laisser aucune place à l'imprévu.* ~ Une fois sur le lit, assis à la droite de celui / qui était désormais son prisonnier, / Dimitri lui *ordonna / d'ôter son uniforme.* ~ Le gardien était resté étrangement silencieux depuis le début de son " enlèvement ", /// il se déshabilla / toujours sans ouvrir les lèvres. ~ *Dimitri s'empara alors du bandeau de tissu / qu'il avait déchiré sur le bas de son* drap la nuit précédente, /// il en entoura la tête du gardien par deux fois au niveau de la bouche. ~ Une fois le bâillon en place, / il prit les menottes / qui pendaient à la

[38] À chaque frontière d'unité, seule l'unité la plus haute dans la hiérarchie est indiquée : « ~ » pour « P » ; « /// » pour « T » ; « // » pour « A » ; « / » pour « C » ; « [ ] » pour les clauses enchâssées.

ceinture de l'uniforme // et attacha le gardien au lit. ~ Il était désormais temps / qu'il quitte à son tour ses vêtements. ~ *Il enfila ensuite ceux / qu'il avait prélevés au* gardien. ~ Avant de passer la porte de ce / qui avait été sa cellule, il veilla / à ce que la casquette soit suffisamment enfoncée sur sa tête / pour que son visage soit en partie recouvert. ~ Il sortit // et regarda sa montre. ~ Dix-huit heures trente, il était *parfaitement à l'heure. ~ C'était l'heure de la r*elève entre gardiens, /// ceux du *service de nuit étaient arrivés, /// il pouvait donc s'en aller.* ~ En saluant vaguement *deux ou trois " collègues ", / il traversa la cour / qui menait à la porte d'entrée.* ~ À dix-huit heures trente-cinq, il était libre. [S11]

Par ailleurs, un des intérêts de notre travail sera d'examiner la variation de durée des pauses en fonction de leur emplacement syntaxique, mais de façon à ce que les emplacements soient mutuellement exclusifs. C'est pourquoi nous avons noté « P » le premier mot d'une phrase ; « T » le premier mot d'une unité T qui ne soit pas une nouvelle phrase ; « A » le premier mot d'une unité A qui ne soit pas une nouvelle unité T, et « C » le premier mot d'une clause qui ne soit pas une nouvelle unité A. Cela est illustré par l'exemple (7).

(7) (P) Le couteau [(C) dérobé le midi même à la cantine] était tenu fermement par sa main droite ; (T) *d'un pas rapide,* il se plaça devant le gardien, (A) lui mit la lame sous la gorge (A) et lui intima (C) *d'être silencieux.*

2.1.2. Codage de l'empaquetage syntaxique

Nous avons indiqué les liens syntaxiques qu'entretient chaque clause avec les clauses qui l'entourent. Il est possible de caractériser l'empaquetage syntaxique à deux niveaux : d'une part, le lien unissant les propositions matrices de deux unités A ; d'autre part, le lien qu'entretiennent les clauses subordonnées par rapport à la clause matrice.

2.1.2.1. Liens entre deux unités A

Juxtaposées

Aucune conjonction de coordination ne lie l'unité A avec la précédente (8).

(8) *il s'entra*inait depuis des mois pour être sûr que son plan fonctionnerait // ***il n'avait pas le droit à l'erreur*** [S08][39]

[39] Le passage en gras correspond au type de clause dont il est question dans la section : par exemple, ici, la clause juxtaposée.

Coordonnées

Les deux clauses sont liées par une conjonction de coordination (hormis « donc », considéré comme un adverbe de liaison), comme dans l'exemple (9).

(9) Alors Dimitri poussa de toutes ses forces // ***et la grille s'arracha*** [S24]

2.1.2.2. Liens entre la clause subordonnée et la clause matrice

Types de clauses subordonnées

Circonstancielles

Les propositions circonstancielles peuvent être finies (10) ou non finies, si elles contiennent un participe passé, un participe présent, un gérondif ou un infinitif, les circonstancielles non finies pouvant avoir leur sujet propre (11).

(10) **quand il prit le couloir qui menait à la sortie**, / il se heurta à Boris [S26]

(11) **ses sensations revenues**, / *il s'engagea* dans le couloir [S18]

Dans certains cas, les indications de lieu ou de temps sont introduites par des expressions composées d'un nom générique, suivi d'un pronom relatif : « au moment où », « là où ». Ces expressions fonctionnant comme des conjonctions de subordination (on peut rapprocher « au moment où » de « lorsque »), nous avons comptabilisé ces clauses parmi les circonstancielles (12).

(12) et *l'abattit sur le crâne du gardien* / **au moment où celui-ci entrait dans sa cellule** [S22]

Relatives

Les clauses relatives prototypiques sont introduites par un pronom relatif, qui est coréférent à un élément de la clause matrice. La forme du pronom relatif varie suivant la fonction syntaxique qu'il occupe dans la clause relative : il peut être sujet (« qui ») (13), objet (« que ») ou oblique (e.g. « avec qui », « dans laquelle », « où », « dont »).

(13) *il l'enferma à son tour dans la cellule* / **qui avait été la sienne pendant dix longs mois** [S28]

Il arrive que l'antécédent soit un générique, par exemple « celui » s'il s'agit d'un humain masculin (14), ou « ce » s'il s'agit d'un référent non-animé. Tout comme les cas de relatives sans antécédent (cas non relevé dans notre corpus), il s'agit de « relatives libres » (Creissels, 2006).

(14) Parmi ces voix Dimitri reconnut celle de celui / ***qui l'avait fait tant souffrir.*** [S30]

En outre, certaines clauses pourraient être considérées comme des circonstancielles ou complétives non finies ; cependant, elles sont parfois considérées comme des relatives non finies (Creissels, 2006), qui pourrait être reformulées par des relatives finies (15-20).

(15) Il devait aller chercher sa fiancée **/ enfermée par le roi elle aussi** [S17]

(16) il devait aller chercher sa fiancée / qui était enfermée par le roi elle aussi

(17) et Dimitri, **/ *craignant d'ameuter les autres gardiens de sécurité*** [S09]

(18) *et Dimitri, / qui craignait d'ameuter les autres gardiens de sécurité*

(19) Dimitri savait bien / que Boris était le seul gardien **/ à pouvoir lutter contre lui** [S26]

(20) Dimitri savait bien / que Boris était le seul gardien / qui pouvait lutter contre lui

Comme les fonctions syntaxiques et sémantiques (telles qu'elles seront définies par la suite) de ces clauses sont similaires à celles des relatives finies, nous les avons comptabilisées parmi les relatives.

Complétives

Les clauses complétives servent à saturer la valence d'un verbe (21), d'un nom (22) ou d'un adjectif (23). Ces clauses peuvent être finies ou non finies (24). Le sujet du verbe non fléchi d'une clause non finie peut être coréférentiel au sujet (23), à l'objet (24), ou à l'oblique de la clause matrice.

(21) Dimitri vit **/ que les étoiles avaient disparu** [S03]

(22) il tentait de garder le fol espoir **/ qu'un jour on les libèrerait** [S27]

(23) il était pressé **/ de quitter la prison** [S14]

(24) *il l'obligea /* **à se taire** [S13]

Parmi les complétives, nous avons également spécifié celles qui étaient des interrogatives, qu'elles soient totales (25) ou partielles (26). Ce type de complétives peut être introduit par un pronom interrogatif (25), l'expression « ce + pronom » (26) ou la conjonction « si » (27).

(25) ne comprenant toujours pas **/ comment il en était arrivé là** [S23]

(26) il lui demanda **/ *ce qu'il s'était passé cette nuit*** [S04]

(27) mais sans savoir **/ *si c'était à cause de l'état ridicule [...]*** [S29]

Par ailleurs, en suivant Creissels (2006), nous avons considéré que les phrases du type (28) étaient constituées d'un verbe régissant deux termes distincts : un syntagme nominal objet et un groupe verbal à tête infinitivale, ce dernier étant codé comme une clause complétive non finie ; les crochets indiquent dans l'exemple suivant les différents syntagmes.

(28) [Il] [entendait] [les pas du gardien] **/ [[ ] [se rapprocher]]** [S25]

Corrélatives

Les clauses corrélatives sont attendues suite à la présence d'un élément de la clause matrice, c'est pourquoi on parle de « corrélation » (29).

(29) Tout avait été si longuement préparé **/ que le plan ne pouvait faillir** [S05]

Spécifications concernant la clause matrice

Indication des clauses matrices

Nous avons explicitement identifié les clauses « matrices », c'est-à-dire celles qui sont liées à des propositions subordonnées. Une clause matrice peut être, elle-même, coordonnée (30), juxtaposée ou subordonnée (31).

(30) **et Dimitri sortit /** (coordonnée matrice)
pour aller diner (circonstancielle) [S10]

(31) **il rit de ce même rire /** (juxtaposée matrice)
qu'entendait Dimitri durant les séances **/** (relative matrice)
où on le torturait (relative) [S30]

Clauses subordonnées enchâssées dans la clause matrice

Les clauses matrices sont généralement suivies, de manière linéaire, par la clause subordonnée. Cependant, dans certains cas la clause subordonnée est enchâssée au sein de la clause matrice (32).

(32) il jeta le livre **/ *qu'il avait soigneusement dissimulé* /** *à l'autre bout de sa cellule* [S02]

Nous avons donc indiqué, au niveau des clauses subordonnées (ici en gras) si elles étaient enchâssées ; et au niveau des clauses matrices, si elles contenaient une clause enchâssée.

Clauses subordonnées précédant la clause matrice

Dans certains cas, la proposition subordonnée se trouve en début d'unité A, puis est suivie de la clause matrice (33).

(33) **Dès que la porte fut ouverte, /** Dimitri bondit sur le gardien [S17]

Nous avons donc indiqué ces cas particuliers où la clause subordonnée précède la clause matrice, et donc où la clause matrice suit la clause subordonnée.

2.1.2.3. Indications sur le discours direct

La présence de dialogue direct crée une rupture dans le récit : d'un point de vue discursif, il y a un changement de situation d'énonciation ; d'un point de vue syntaxique, il arrive que le discours direct vienne saturer la valence de certains verbes, mais ce n'est pas toujours le cas. C'est pourquoi nous avons spécifié ces différentes clauses : discours direct (34), avant discours direct (35), après discours direct (36).

(34) **« Sors, Dimitri ! »** [S23]

(35) **Celui-ci lui cria :** / « Reviens ici / ou nous te suivrons à vie ! » [S10]

(36) « Donne-moi les clés » / **lui murmura-t-*il à l'oreille*** [S03]

2.1.2.4. Résumé du codage syntaxique des clauses

Les informations liées à l'empaquetage syntaxique qui seront utilisées lors des analyses ultérieures sont résumées dans le Tableau 4.

Type de clause	**Caractéristique de la subordonnée**	**Position par rapport aux autres clauses**		**Remarque liée au discours direct**
		Matrice	**Subordonnée**	
- juxtaposée - coordonnée - circonstancielle - complétive - relative - corrélative	- non finie - finie - interrogative indirecte	- matrice -matrice contenant une enchâssée - juxtaposée après circonstancielle	- enchâssée - circonstancielle avant clause matrice	- discours direct - avant discours direct - après discours direct

Tableau 4. Codage syntaxique des clauses – Résumé

2.2. Données chronométriques

2.2.1. Adaptation des données chronométriques pour un texte

Notre étude porte sur des textes, à différents niveaux d'unités, ce qui nous a amenés à effectuer quelques adaptations.

2.2.1.1. « Pauses initiales » et ponctuation

La présence de signes de ponctuation nous a interrogés sur ce que nous appelions « pause initiale » : la pause précédant une phrase débute-t-elle avant ou après le point ? Si deux clauses sont séparées par une virgule, la pause initiale de la seconde clause débute-t-elle avant ou après la virgule ?

La fonction de la pause avant et après le signe de ponctuation n'est pas définie de manière exacte, même si intuitivement on pourrait penser que la pause avant la ponctuation servirait plutôt à la révision, et celle après servirait plutôt à la planification. Cependant, cette hypothèse peut être soumise à des différences interindividuelles et intraindividuelles (Foulin et al., 1989). En outre, dans le cas de deux clauses qui ne seraient pas séparées par un signe de ponctuation, les deux fonctions distinguées ci-dessus – révision et planification – seraient assignées à une même pause.

Nous avons donc choisi de considérer comme « pause initiale » la somme de la pause initiale du signe de ponctuation et de la pause initiale du premier mot de l'unité. Nous avons cependant indiqué les cas où une révision était effectuée entre la ponctuation et le premier mot, cela pouvant rallonger fortement le temps de pause.

2.2.1.2. Retours à la ligne

Les textes faisant souvent environ une page, il s'est posé le problème des retours à la ligne : beaucoup de clauses s'étendaient sur plus d'une ligne, mais la pause précédant le premier mot de la ligne pouvait être beaucoup plus longue uniquement à cause du déplacement du stylo. Afin de minimiser cet effet, nous avons relevé pour chaque participant la pause la plus courte devant un mot écrit après un retour à la ligne, à laquelle nous avons retranché la durée de la pause la plus courte dans le reste du texte : la durée obtenue a été considérée comme le temps nécessaire « physiquement », pour ce scripteur, pour passer d'une ligne à l'autre. Nous avons ensuite soustrait ce temps à chaque pause initiale de mot situé en début de ligne.

2.2.2. Données chronométriques pour chaque unité

Les variables chronométriques ont été associées aux unités Telles que segmentées dans notre texte. Outre les pauses initiales, chaque unité est caractérisée par la temporalité de son écriture interne (comprenant les pauses inter-mots, les pauses intra-mots et les temps d'exécution).

2.2.2.1. Pauses initiales

Tout d'abord, pour chaque unité (phrase, unité T, unité A, clause), la variable « pause initiale » a été indiquée. Ainsi, la pause initiale pouvait, grâce au codage, être associée à la pause précédant une phrase, une unité T, une unité A ou une unité C. Par ailleurs, un des intérêts de notre travail sera de voir la distinction de durée entre les pauses en fonction de l'emplacement syntaxique, mais de façon à ce que les emplacements soient mutuellement exclusifs. C'est pourquoi, grâce aux codages « P », « T », « A »,

« C », il sera possible de connaitre par exemple la durée des pauses en initiale de phrase, celles en initiale d'unité T (début de nouvelle phrase compris), mais également celles en initiale d'unité T qui ne soit pas une nouvelle phrase (« T »).

2.2.2.2. Moyenne des « pauses intra-mots » de l'unité

Il s'agit de la somme des pauses internes de chaque séquence de l'unité, divisée par le nombre de caractères de l'unité. Nous avons ainsi la moyennes des pauses intra-mots de chaque phrase, chaque unité T, chaque unité A et chaque clause.

2.2.2.3. Moyenne des « pauses inter-mots » de l'unité

Il s'agit de la somme des pauses avant chaque séquence de l'unité, en excluant la pause devant le premier mot, divisée par le nombre de séquences de l'unité moins une séquence. Il a été choisi d'exclure la pause initiale, puisque celle-ci pouvait faire l'objet d'un traitement indépendant grâce à la variable « pause initiale de l'unité ». Nous avons ainsi la moyenne des pauses inter-mots de chaque phrase, chaque unité T, chaque unité A et chaque clause.

2.2.2.4. Moyenne du temps d'exécution

Le temps d'exécution correspond à la durée d'écriture de l'unité, après avoir enlevé toutes les pauses, divisée par le nombre de caractères de l'unité. Nous avons ainsi la moyenne du temps d'exécution de chaque phrase, chaque unité T, chaque unité A et chaque clause.

2.2.2.5. Moyenne de la durée d'écriture interne de l'unité

Il s'agit de la durée d'écriture de l'unité, après avoir seulement enlevé la pause initiale. Il s'agit donc de la somme des pauses inter-mots, pauses intra-mots et temps d'exécution de l'unité, divisée par le nombre de séquences de l'unité. Nous avons ainsi le temps d'écriture interne[40] de chaque phrase, chaque unité T, chaque unité A et chaque clause.

2.2.3. Utilisation des déciles et des quartiles

Nous avons généralement utilisé les moyennes des différentes données, ou des outils statistiques permettant de conserver les données tout en prenant en compte la variable « participant ». Cependant, dans une perspective de comparaison entre individus, il n'est pas toujours pertinent de raisonner en termes de comparaison de moyennes, les seuils pouvant différer d'un participant à l'autre.

[40] L' « écriture interne » comprend donc les temps de pause et d'exécution, à l'exception de la pause initiale de l'unité.

Pour certaines analyses, nous avons donc eu recours à la notion de quartiles ou de déciles. Cela permet de répartir les pauses en fonction de leur longueur relativement aux autres pauses du texte. Pour cela, les pauses de chaque texte ont été classées par ordre croissant. Les quartiles partagent en quatre parties d'effectifs égaux les données, alors que les déciles partagent les données en dix parties d'effectifs égaux. Nous avons nommé « Q_i » (ou « D_i ») les pauses dont la durée est comprise entre Q_{i-1} (ou D_{i-1}) et Q_i (ou D_i). Par exemple, les pauses « Q1 » correspondent aux 25% de pauses les plus courtes du texte, alors que les pauses « D10 » correspondent aux 10% de pauses les plus longues du texte. Cela permet par exemple de voir si, de manière comparable entre tous les individus, les pauses en initiales de phrase se situent plutôt parmi les 25% de pauses les plus longues (Q4), plus précisément parmi les 10% de pauses les plus longues (D10), ou si elles sont parmi les pauses plus courtes.

3. Analyses utilisées

Plusieurs types d'analyse ont été utilisés dans les études menées pour cette thèse. Les analyses varient suivant la finalité de l'étude, le type de données à notre disposition, et l'avancement de notre réflexion méthodologique au moment de l'étude. Ces analyses ont été menées grâce au logiciel SPSS. Pour chaque étude il sera précisé quelles analyses ont été faites, mais nous donnons ici quelques éléments généraux.

Tout d'abord, il convient de préciser que nos variables dépendantes étant des données chronométriques, leur distribution présente une forte asymétrie à droite. Afin de normaliser ces données, pour pouvoir effectuer certains types d'analyses et obtenir un modèle plus ajusté, nous avons utilisé la transformation « log ». Cependant, afin d'illustrer les résultats de manière plus concrète, les moyennes présentées seront données en millisecondes (ms).

3.1. Explication de la variance de données continues

3.1.1. Analyse de Variance (ANOVA)

Des ANOVAs ont été réalisées pour voir dans quelle mesure des variables dépendantes (VD), telle que la durée des pauses, différaient en fonction de certains groupes. Ces groupes pouvaient être de deux types. D'une part, il pouvait s'agir de groupes de participants différents. C'est notamment le cas dans les études développementales : une ANOVA à un facteur comprenant 3 modalités peut être réalisée pour voir par exemple dans quelle mesure la durée des pauses initiales varie en fonction du niveau scolaire des participants. Il est également possible de procéder à des analyses de variance avec plusieurs variables indépendantes (VI) : par exemple, voir dans quelle mesure la durée des pauses initiales varie en fonction du niveau scolaire des participants et de l'ordre de production, en regardant s'il y a un effet d'interaction entre ces deux VI. Dans ces cas, les données sont moyennées pour chaque texte, puis l'analyse est menée sur ces données. D'autre part, la VI pouvait correspondre à une variable linguistique : par exemple, l'emplacement syntaxique (début de phrase, de clause, etc.) ou un type de clause particulier (e.g. juxtaposée, complétive).

S'il y avait suffisamment de données disponibles, l'ANOVA était effectuée sur les moyennes par texte ; sinon, l'analyse était menée sur l'ensemble des données. Dans le cas où un facteur à plus de deux modalités avait un effet significatif, nous avons mené des tests post-hoc entre chaque paire grâce au test de Bonferroni.

3.1.2. Régression

Dans certains cas, nous avons souhaité voir dans quelle mesure plusieurs VI, qu'elles soient continues ou catégorielles, pouvaient expliquer la variance d'une VD continue : par exemple, dans

quelle mesure le temps d'exécution d'une clause dépend de sa longueur et de sa position dans la phrase. Nous avons alors eu recours à des régressions multiples. L'équation linéaire d'une régression multiple peut être exprimée sous la forme :

$$Y = a + b_1*X_1 + b_2*X_2 + \ldots + b_k*X_k + \varepsilon$$

Dans cette équation, Y correspond à la VD, « a » correspond à la constante des effets fixes, « b_i » est le coefficient de régression de la variable X_i et ε est l'erreur résiduelle, c'est-à-dire ce qui n'est pas expliqué par les variables introduites.

Afin de voir quelles VI ont un effet significatif, nous avons généralement utilisé la méthode « pas-à-pas », générée automatiquement par SPSS : la variable ayant la plus haute corrélation avec la VD est introduite en premier, puis c'est la variable qui a la plus forte corrélation partielle avec la VD (c'est-à-dire, la plus forte corrélation par rapport à la variance qu'il reste à expliquer), et ce jusqu'à ce que l'ajout de VI ne conduise pas à un apport significatif par rapport au modèle précédent. Les variables redondantes sont en outre éliminées. Ce type d'analyse nécessite que les VI introduites ne soient pas corrélées de manière trop élevée entre elles : nous nous sommes donc assurés qu'aucune corrélation n'avait un coefficient supérieur à 0,7.

Deux résultats peuvent particulièrement être intéressants à l'issue de ces analyses : la part de variance expliquée (R^2), par le modèle complet et par chaque variable introduite, ainsi que le signe du coefficient bêta, qui indique, dans le cas d'une VI continue, si la relation entre la VI et la VD est positive (lorsque la VI augmente, la VD augmente) ou négative (lorsque la VI augmente, la VD diminue). Généralement, les analyses étaient menées sur l'ensemble des données.

3.1.3. Modèles linéaires mixtes

Nous avons vu qu'il était possible d'expliquer la variance d'une VD en ayant soit recours à des ANOVAs, soit à des modèles de régressions multiples. Cependant, ces tests ne reflètent pas toujours de manière exacte la structure de nos données, qui sont des unités linguistiques écrites par des participants : une des difficultés pour voir l'effet de l'emplacement syntaxique sur la durée de la pause initiale est de prendre en compte la variabilité interindividuelle tout en évitant de trop simplifier les données, au risque de perdre de la précision. En effet, si l'on fait une simple ANOVA, il faut soit utiliser l'ensemble des données, ce qui masque le fait que ces données ne sont pas indépendantes, puisque plusieurs données sont produites par un même individu ; soit moyenner les données par emplacement syntaxique et par sujet, mais au risque de masquer la variabilité des données, et d'accorder une trop grande place aux valeurs extrêmes. De même, dans le cas des analyses de régressions multiples, lorsque nous utilisons l'ensemble des données, nous perdons l'information concernant le participant.

La meilleure solution pour ce type de données est alors d'utiliser des modèles linéaires mixtes, qui correspondent bien à la structure de nombreuses données de psycholinguistique (Baayen, Davidson &

Bates, 2008). Ce type de modèle, appelé également « multiniveaux », permet de prendre en compte simultanément une structure hiérarchisée : par exemple, des individus produisant des mots. On peut en outre spécifier quels sont les effets fixes et quels sont les effets aléatoires. Les effets fixes sont dus à des variables dont les différentes modalités représentent l'ensemble des possibilités, et sont choisies par l'expérimentateur. Par exemple, si nous regardons dans quelle mesure la pause initiale varie en fonction de l'emplacement syntaxique, en choisissant quatre niveaux d'emplacement syntaxiques, ces quatre emplacements syntaxiques correspondent aux quatre modalités d'un effet fixe. Les effets aléatoires, quant à eux, sont dus à des facteurs qui ne sont qu'un échantillon de l'ensemble de la population, et ne sont pas réellement choisis par l'expérimentateur. Dans le cas de nos études, la variable « participant » est un effet aléatoire, puisque tous les cas de variabilité inter-participants ne sont pas pris en compte. Le but du modèle linéaire mixte est alors d'expliquer la variance d'une VD en fonction d'effets fixes et d'effets aléatoires.

Le modèle linéaire mixte correspond à une régression multiple, dans laquelle sont introduits des effets fixes et des effets aléatoires. On peut exprimer l'équation linéaire sous la forme :

$$Y = a + b_1*X_1 + b_2*X_2 + \ldots + b_k*X_k + \varepsilon$$

Là encore, Y correspond à la VD, « a » correspond à la constante des effets fixes, « b_i » est le coefficient de régression de la variable X_i et ε est l'erreur résiduelle. La différence avec la régression multiple « simple » est que la variable X_i peut être soit un effet fixe soit un effet aléatoire.

La procédure pour sélectionner le meilleur modèle issu d'un modèle linéaire mixte est moins automatisée que pour les régressions multiples (où le logiciel SPSS sélectionne ou élimine lui-même les variables). Le recours à des modèles mixtes nécessite donc plusieurs étapes afin d'obtenir les résultats les plus pertinents.

Tout d'abord, il faut établir un « modèle vide » (« modèle 0 »), dans lequel seuls 3 paramètres sont insérés : la constante des effets fixes (« a » dans l'équation ci-dessus), la variation inter-participants (une variable « X » qui correspond au facteur aléatoire « participant ») et la variation intra-participant, qui correspond au résidu (« ε » dans l'équation ci-dessus). Grâce au modèle vide, on peut calculer la proportion de variance inter-participants par rapport à la variance totale. Pour cela, on utilise l'équation suivante :

$$ICC = \frac{variance\ du\ groupe\ \text{"participants"}}{variance\ résiduelle + variance\ du\ groupe\ \text{"}participants\text{"}}$$

Le résultat de ce calcul est appelé « ICC » (Intraclass Correlation Coefficient) et peut être exprimé en pourcentage. Un ICC élevé indique qu'une part importante de la variance totale est due aux différences entre participants. Si l'ICC est quasi-nul, il n'est pas justifié d'utiliser un modèle linéaire mixte.

Nous insérons ensuite les effets fixes qui, selon nos hypothèses, sont susceptibles d'influencer la variable dépendante. Il n'est pas possible d'insérer simultanément plusieurs effets fixes trop corrélés (si $r > 0{,}70$), au risque de biaiser les résultats.

L'intérêt est finalement d'obtenir le modèle le plus parcimonieux possible, c'est-à-dire celui qui ajuste au mieux les données avec le moins de paramètres possible. Afin de vérifier que le modèle n ajuste mieux les données que le modèle précédent n-1, nous avons utilisé la décroissance de la déviance, qui compare la vraisemblance de chaque modèle grâce à l'équation :

$$\Delta = (-2 \log L_{\text{modèle n-1}}) - (-2 \log L_{\text{modèle n}})$$

Grâce à une table de χ^2, nous vérifions que cette différence est significative en fonction du nombre de paramètres. Par exemple, si $\Delta = 6{,}3$ alors que le modèle 1 n'a qu'un paramètre supplémentaire (par exemple, un effet fixe à 2 modalités), la différence est significative et le modèle est donc meilleur ; en revanche, si $\Delta = 6{,}3$ mais que 4 paramètres ont été ajoutés au modèle 1 (par exemple, 2 effets fixes à 3 modalités chacun), la différence entre les deux modèles n'est pas significative. Afin de réduire le nombre de paramètres, nous avons parfois regroupé certaines modalités. Cela a été fait grâce à l'observation des tests post-hoc de Bonferroni, qui permettent de révéler quelles sont les différences significatives entre les modalités de chaque facteur, et donc quelles modalités pourraient être regroupées, tout en s'assurant que ces regroupements sont justifiés d'un point de vue théorique.

Afin de savoir dans quelle mesure les effets fixes introduits dans un nouveau modèle l'améliorent par rapport à l'ancien, nous indiquons le « pseudo R^2 ». Ce gain explicatif correspond à la différence de l'erreur résiduelle entre les deux modèles par rapport à l'erreur résiduelle du premier modèle, et est exprimé en pourcentage. Il se calcule grâce à l'équation :

$$Pseudo\ R^2 = \frac{(résidu\ du\ modèle\ n-1) - (résidu\ du\ modèle\ n)}{(résidu\ du\ modèle\ n-1)}$$

Dans les analyses menées, nous ne mentionnons parfois que le modèle final, afin de présenter directement des résultats clairs. Lorsque peu de VI sont introduites au départ, nous considérons comme « modèle final » le modèle le plus parcimonieux, c'est-à-dire le modèle pour lequel l'ajout de paramètres n'améliorerait pas de manière significative la statistique de la déviance. Dans le cas où de nombreuses VI sont introduites au départ, il s'agira du modèle qui est meilleur que le modèle 0 tout en gardant le maximum de VI.

3.2. Établissement de profils

Au cours de nos études, nous avons parfois voulu établir des « profils » de scripteurs. En effet, même au sein d'un groupe qui pourrait paraitre homogène, il peut y avoir des différences interindividuelles. Or, il est intéressant de les mettre à jour, afin notamment d'expliciter par rapport à quelles variables on peut définir des types de scripteurs différents. En outre, cette problématique de « profils » est

particulièrement pertinente d'un point de vue développemental, où les différences de performance entre élèves sont souvent relevées. Nous avons donc établi des profils dans une des études développementales, en lien avec la maitrise de la ponctuation, et dans la description générale des données chronométriques des adultes.

La première étape consiste à classer les participants en différents « profils » : c'est ce que permettent les analyses en nuées dynamiques et les classifications hiérarchiques. La deuxième étape permet de voir quelles variables sont significativement différentes en fonction des groupes ; pour cela nous avons utilisé les analyses discriminantes.

3.2.1. Création de classes de profils

Pour créer les classes de profils, il faut tout d'abord sélectionner les variables à entrer dans l'analyse. Ce choix est en premier lieu théorique, puisqu'il faut se demander à partir de quelles données on veut observer le comportement des scripteurs. Par exemple, dans la description générale on observera plusieurs comportements, comme la durée absolue des données chronométriques, leur répartition, les changements au cours de l'écriture du texte, etc. Dans le cas où deux variables étaient corrélées à plus de 0,7, seule une des variables a été gardée, afin de conserver l'indépendance des données, tout en évitant de perdre trop d'informations en écartant un nombre important de données.

Une première manière d'établir des profils est d'utiliser le test « nuée dynamique » (ou k-means clusters). Un des inconvénients de ce test est qu'il faut indiquer le nombre de classes que l'on souhaite obtenir au final. Cependant, nous n'avions pas a priori de nombre de classes défini, hormis le fait que, pour des résultats qui restent pertinents, il faut qu'il y ait entre 2 et 4 groupes. Après avoir inséré les variables, nous avons donc comparé les répartitions en 2, 3 et 4 groupes. L'observation graphique des résultats (établie grâce aux centres finaux des classes obtenues) ainsi que du nombre de scripteurs par profil a permis de dégager le nombre de profils à garder le plus pertinent.

Il existe une autre possibilité sous SPSS pour établir des classes : le test « classification hiérarchique ». Nous avons utilisé la méthode de Ward, à partir du carré des distances euclidiennes ; les variables continues ont été centrées-réduites. Le résultat des classifications hiérarchiques se présente sous forme d'un arbre : plus les « branches » sont longues, plus la distance entre les classes est importante ; à l'inverse, deux individus graphiquement côte à côte et reliés par une branche très courte ont un comportement très similaire. Afin de déterminer le nombre de classes, plusieurs critères ont été pris en compte : la distance entre les classes (nous avons plutôt segmenté au niveau de branches « longues »), le nombre de cas par classe, et l'interprétation possible des résultats. Généralement, entre 2 et 4 groupes ont été retenus pour chaque profil. La Figure 11 illustre le type de résultat obtenu grâce au logiciel. Le trait en pointillé indique le niveau où nous avons segmenté, et les cadres en gras font ressortir les différents groupes de scripteurs (indentifiables grâce à leur numéro, de 1 à 30). Ici, nous avons donc 3 profils. L'arbre indique que le « profil 3 », composé de seulement 5 scripteurs, diffère

fortement des autres. Les individus adjacents et reliés par une branche très courte, comme les participants S18 et S20, ont un comportement très proche.

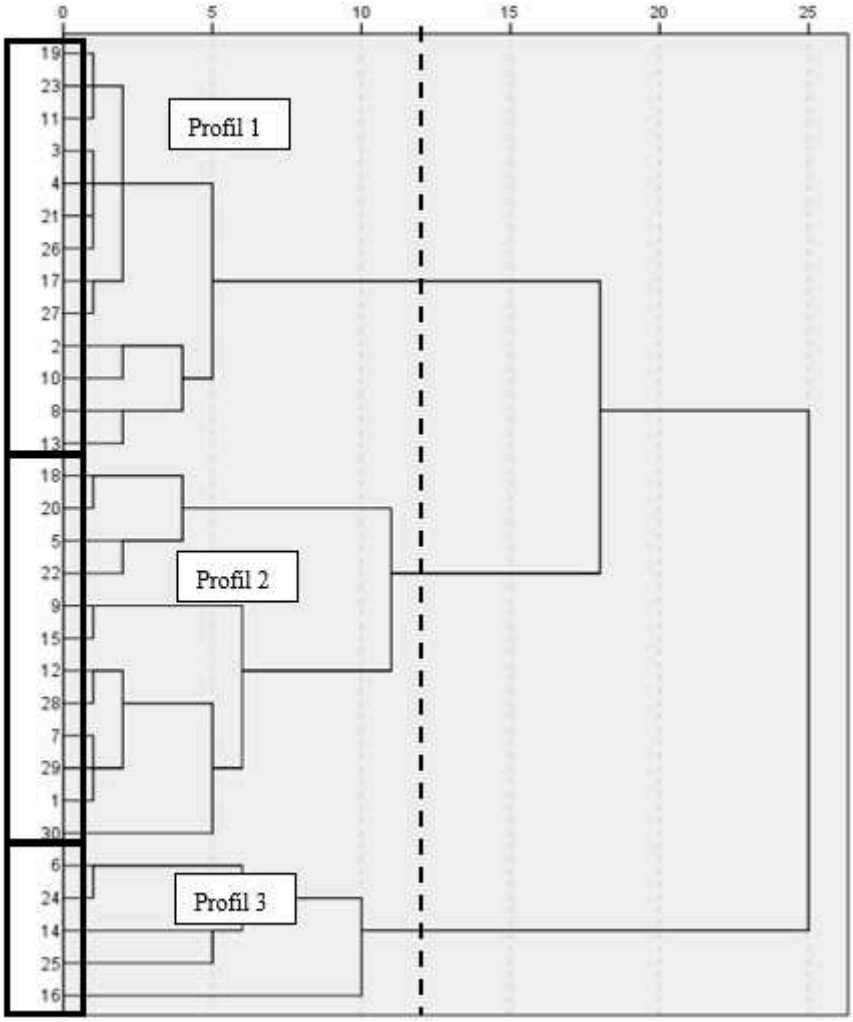


Figure 11. Exemple de profils établis par la classification hiérarchique

3.2.2. Détermination des variables significativement différentes

Les deux méthodes présentées ci-dessus donnent le même type de résultat : un nombre de profils, avec pour chaque variable la moyenne pour chaque profil. Pour savoir dans quelle mesure ces profils étaient robustes et comment les caractériser, nous avons ensuite utilisé l'analyse discriminante. Pour cela, nous avons indiqué tout d'abord une variable catégorielle (ici le profil établi). Nous avons ensuite entré les variables qui avaient permis de créer les profils. Cela permettait de « reconstituer » les classes, et de manière logique les participants étaient parfaitement classés (généralement 100% d'affectation correcte). Avec SPSS, il est possible d'obtenir des résultats d'ANOVAs indiquant quelles variables diffèrent significativement d'un profil à l'autre. Ainsi, nous pouvons d'une part donner pour chaque profil la moyenne de chaque variable, et d'autre part indiquer quelles variables permettent de distinguer les profils obtenus.

Résultats

Chapitre 5 – Les unités de production : aspect développemental

Une des questions principales de notre travail concerne les unités de production : quel est l'empan de la micro-planification et / ou de la formulation ? Les études précédentes ont suggéré que l'empan de production des enfants serait davantage la clause que la phrase (Chanquoy et al., 1990; Foulin, 1998; van Hell et al., 2008). Ces études portaient principalement sur des enfants entre 7 et 10 ans. Ce plus faible empan de planification peut tout d'abord s'expliquer par des contraintes cognitives : à cet âge, l'exécution motrice n'est pas encore automatisée (Berninger & Swanson, 1994) et les ressources cognitives n'ont pas atteint leur capacité maximale (Gathercole et al., 2004). Les difficultés liées à la transcription rendent l'exécution plus lente et couteuse, ce qui peut empêcher une planification sur un trop grand segment de texte, surtout si les capacités mnésiques sont moins élevées. En outre, la conformité aux normes de ponctuation n'est pas encore acquise en 6ème (Fauchart, 2001), ce qui suggère que pour les élèves de primaire la notion de « phrase » n'est pas tout à fait comprise. Or, même si certaines études indiquent que les enfants ne se comportent pas comme les adultes quant à la gestion temporelle des clauses et des phrases, aucune ne présente la manière dont ce comportement se développe jusqu'à se rapprocher de celui des adultes. Il semblerait alors pertinent de regarder quelles sont les unités de planification[41] d'élèves ayant entre 10 et 15 ans : alors qu'à 10 ans l'exécution motrice n'est pas tout à fait automatisée et des problèmes de ponctuation persistent, à 15 ans ces compétences ont fortement évolué, même si elles peuvent encore gagner en expertise. Cette tranche d'âge pourrait donc permettre de faire ressortir le pattern développemental lié à l'empan de planification.

Ayant à notre disposition le corpus Reflex, un corpus développemental consistant en des textes rédigés par des élèves de 10 à 15 ans dont les données chronométriques étaient également enregistrées, nous avions le matériau nécessaire pour répondre à la question des unités de production dans une perspective développementale. Nous avons donc appliqué à ce corpus le codage décrit précédemment, afin de recourir à plusieurs indices permettant d'identifier les unités de production. D'un point de vue linguistique, la segmentation en unités T permet d'avoir une unité intermédiaire entre la clause et la phrase ; l'indication de l'empaquetage syntaxique permettra d'analyser la gestion temporelle de la production de clauses plus ou moins dépendantes d'un point de vue syntaxique. D'un point de vue chronométrique, le recours à la durée des temps d'exécution permettra d'affiner les résultats obtenus à partir de l'étude des pauses.

[41] Le terme « planification » est utilisé dans notre travail par facilité : il désigne de manière plus précise le fait d'effectuer des choix sur ce qui suit, que ce soit au niveau sémantique, syntaxique et lexical – ce qui regroupe les notions de « micro-planification » et « formulation ».

Dans un premier temps, nous interrogerons la question de l'unité de production d'un point de vue développemental. Dans un deuxième temps, nous verrons dans quelle mesure le fait de prendre en compte la maitrise des normes typographiques liées à la ponctuation peut permettre d'affiner ce pattern développemental. Les deux études présentées dans ce chapitre ont été publiées ou sont à paraitre[42]. Pour chaque étude une simple introduction sera fournie pour en présenter le but, les éléments pertinents du cadre théorique ayant été présentés dans les premiers chapitres de cette thèse. Il sera également possible de se reporter au chapitre « Méthodologie générale » pour obtenir plus de détails sur le protocole de collecte de données et les analyses utilisées.

[42] Cela explique certains choix de codage ou de perspective effectués : le codage syntaxique de la première étude correspond à ce qui a été majoritairement utilisé dans le cadre du projet Spencer « Developing literacy in difference contexts and different languages », auquel ont participé beaucoup de contributeurs de l'ouvrage dans lequel elle a été publiée ; la seconde étude aborde davantage des considérations didactiques, puisqu'elle est publiée dans la revue LIDIL.

1. Une perspective développementale des unités de production[43]

1.1. Introduction

Cette étude cherche à identifier les unités de production en mettant en relation les données syntaxiques, telles que les frontières d'unités syntaxiques ou le type de relation entre les clauses, avec les données chronométriques enregistrées, plus particulièrement les pauses initiales et le temps d'exécution. Cette méthodologie s'appuie sur le postulat que les pauses reflètent le temps que prend le scripteur pour planifier et prendre des décisions sur ce qui suit (Matsuhashi, 1982). Des études similaires ont suggéré que la planification pouvait être effectuée sur unité Au-delà de la clause (Immonen & Mäkisalo, 2010; van Hell et a1., 2008). Cependant, il n'est pas certain que les scripteurs possèdent tous cette faculté, et dans toutes les situations : par exemple, les Finnois traduisant un texte anglais planifiaient davantage clause par clause que sur des unités supérieures à la clause, ce qui pouvait être expliqué par le cout de la formulation lors d'une telle activité (Immonen & Mäkisalo, 2010). Il peut également arriver que les ressources dévolues à la micro-planification et à la formulation soient restreintes à cause d'une macro-planification plus couteuse suivant le type de texte, qui peut requérir plutôt une stratégie des connaissances rapportées, comme pour les textes narratifs, ou plutôt une stratégie des connaissances transformées, comme pour les textes expositifs (Bereiter & Scardamalia, 1987).

Nous examinerons ici comment les enfants de différents âges planifient un texte multi-clausal. Outre la sélection des items lexicaux et grammaticaux corrects, les scripteurs français doivent être attentifs à l'orthographe des mots, ce système orthographique étant relativement opaque. La gestion de ces difficultés pourrait les empêcher d'être en mesure de planifier des constructions syntaxiques dépendantes de la même manière que des adultes rédigeant dans leur langue maternelle. Il sera ainsi possible de voir comment la capacité à planifier des unités multi-clausales se développe dans le cadre de l'écriture de textes expositifs et narratifs produits par des élèves de CM2, 5ème et 3ème.

[43] Cette étude reprend, en le traduisant, le chapitre d'ouvrage : Ailhaud, E., Chenu, F. & Jisa, H. (2015). A developmental perspective on the units of written French. In J. Perera, M. Aparici, E. Rosado & N. Salas (Eds.), Écrit and spoken language development across the lifespan: Essays in honour of Liliana Tolchinsky (p. 287–305). Springer.

1.2. Méthodologie

1.2.1. Données utilisées

Les textes étudiés sont issus du corpus Reflex. Les scripteurs, âgés de 10 à 15 ans, étaient répartis en trois groupes de quarante participants, suivant le niveau scolaire (CM2, 5ème, 3ème.). Deux types de textes étaient produits – un expositif et un narratif – sur le thème des problèmes entre les gens. Bien que chaque texte ait été produit à l'écrit et à l'oral, seuls les textes écrits sont utilisés pour cette étude ; il est cependant possible de prendre en compte l'ordre de production (écrit ou oral en premier). Ces textes étaient rédigés sur une tablette à digitaliser et enregistrés grâce au logiciel Eye & Pen.

1.2.2. Codage des données

1.2.2.1. Segmentation du texte

Conformément à la segmentation explicitée dans la méthodologie générale, les textes ont été divisés en clauses, unités Terminales et phrases graphiques. Ainsi, chaque mot a été codé suivant son emplacement : dans cette étude, « P » désigne le premier d'une phrase graphique, « T » désigne le premier mot d'une unité T et « C » désigne le premier mot d'une clause.

1.2.2.2. Caractéristiques syntaxiques

Chaque clause a été codée suivant ses caractéristiques syntaxiques et ses relations avec les clauses qui l'entourent. La « complexité syntaxique » désigne davantage un continuum qu'une opposition binaire entre « structures complexes » et « structures non complexes ». Ce continuum de complexité a fait l'objet de nombreuses études en linguistique générale (Haiman & Thompson, 1988; Hopper & Thompson, 2008; Matthiessen & Thompson, 1988) et plus particulièrement en psycholinguistique, avec une longue tradition d'études sur la compréhension des structures s'étendant sur ce continuum (Pickering & van Gompel, 2006 pour une revue de la littérature). Dans le cadre de notre étude, le type de clause a été codé de manière à rendre compte de ce continuum de complexité syntaxique à partir des notions d'isotaxe, parataxe, hypotaxe et endotaxe (Aparaci, 2010; Berman & Nir, 2009; Nir, 2008; Nir & Berman, 2010), utilisées dans plusieurs études monolinguistiques et translinguistiques issues du projet Spencer[44] étudiant la connectivité syntaxique de textes narratifs et expositifs dans les modalités orales et écrites.

[44] Ce projet regroupait sept pays, l'Espagne, la France, l'Islande, Israël, les Pays-Bas, la Suède et les États-Unis d'Amérique, et , était financée par la fondation Spencer (USA). Le responsable scientifique de ce projet, intitulé « Developing literacy in différent contexts and different languages » était Ruth Berman. Le protocole expérimental (type de tâches demandé) a été repris lors du projet ReFlex.

Isotaxe (ISO)

Cette catégorie inclut des relations de coordination et de juxtaposition (37). Deux clauses sont reliées par l'utilisation d'une conjonction de coordination ou d'autres adverbes de liaison, ou sont simplement juxtaposées.

(37) *Ma sœur était pas loin, /* <u>elle a entendu la conversation</u> [3ème,10,f,n,écrit1][45]

Les deux clauses sont ainsi plutôt indépendantes d'un point de vue syntaxique.

Parataxe (PARA)

Cette catégorie inclut des relations de coordination et de juxtaposition comportant une ellipse du sujet (38) ainsi que les clauses complétives (39).

(38) Une jeune demoiselle derrière elle passa / <u>et le ramassa</u> [5ème,25,m,e,écrit2]

(39) Je pense / <u>que je ne le ferai plus</u>. [3ème,12,m,n,écrit2]

Conformément à la proposition de Nir et Berman (2010), nous considérons les clauses complétives comme de la parataxe asymétrique. En effet, la clause complétive étant requise par le prédicat, la clause matrice et la clause subordonnée ne peuvent pas apparaitre en isolation.

Hypotaxe (HYPO)

Les structures d'hypotaxe comprennent deux catégories : les subordonnées finies et non finies. Les subordonnées finies incluent des clauses subordonnées dont le verbe est fléchi pour le temps et la personne ; il s'agit principalement de clauses relatives (40) et de clauses circonstancielles (41) .

(40) *J'ai raconté une histoire / qui m'arrive souvent avec mes copines* [CM2,40,f,n,écrit2]

(41) *J'allais y aller / quand soudain la femme d'accueil me surprit.* [G7,22,m,n,écrit2]

Les subordonnées non finies incluent des clauses dont le verbe est un infinitif (42), un gérondif (43) ou une forme participiale ; le verbe ne donne donc pas d'indication de temps ou de personne.

(42) Et puis il part / <u>sans nous répondre</u>. [CM2,44,m,n,écrit2]

[45] Les exemples sont identifiés grâce aux éléments suivants : niveau scolaire (CM2, 5ème, 3ème), numéro du sujet, sexe (f=féminin, m=masculin), type de texte (n=narratif, e=expositif), ordre de production (écrit1=écrit en premier, écrit2=écrit en second). La clause dont il est question dans la description du codage est soulignée dans l'exemple.

(43) En revenant de la voile / je me suis disputée avec une camarade de classe [CM2,20 ,f,n,écrit2]

Endotaxe (ENDO)

Ce dernier type de clause concerne les clauses enchâssées (44), c'est-à-dire les clauses subordonnées qui interrompent l'écriture de la clause matrice.

(44) Et mon père / qui veut toujours avoir raison / *s'est emporté* [CM2,19,f,n,écrit1]

Clauses matrices et non matrices (MAT et NMAT)

La distinction entre clauses matrices et non matrices a également été prise en compte. Les clauses matrices sont accompagnées de clauses dépendantes, contrairement aux clauses non matrices (46). La caractéristique « matrice » ou « non matrice » s'applique aux différentes catégories présentées ci-dessus : ainsi, l'exemple suivant comporte deux clauses matrices (une isotaxique et une parataxique) et une clause parataxique non matrice.

(45) En fait je pense [ISO MAT] / *qu'elle ne savait pas* [PARA_MAT] / que le soir nous allions sortir [PARA_NMAT] [G9,26,f,n,écrit2]

Position de la clause dépendante

Enfin, une indication permet d'identifier une structure spécifique concernant la position de la clause subordonnée : il a été indiqué lorsque la clause subordonnée précède la clause matrice (46).

(46) En revenant de la voile [HYPO avant] / je me suis disputée avec une camarade de classe [ISO après] [CM2,20 ,f,n,écrit2]

1.2.3. Analyses

Nous avons principalement mené des analyses de variance (ANOVAs). Celles-ci prennent en compte trois variables indépendantes (VI) : le niveau scolaire (CM2, 5ème, 3ème), le type de texte (narratif, expositif) et l'ordre de production (oral-écrit, écrit-oral). Les variables dépendantes (VD) sont les pauses initiales (en millisecondes) et le temps d'exécution (en millisecondes par caractère). Afin de pouvoir mener les analyses statistiques, les données ont été normalisées grâce à une transformation en log. Cependant, les nombres indiqués dans les figures et le texte correspondent aux données non transformées.

1.2.4. Hypothèses

Plusieurs hypothèses peuvent être posées concernant la variation des données chronométriques en fonction des VI.

Concernant l'ordre de production, il est attendu que l'amorçage provoqué par la production orale conduise à une production plus rapide des textes « écrit2 », le contenu ayant pu être en partie planifié durant la production orale. Les pauses initiales et le temps d'exécution devraient donc être plus courts pour les textes « écrit2 », par rapport aux textes « écrit1 ».

Le type de texte devrait également affecter les processus cognitifs. Les textes narratifs, produits plus précocement par les enfants et ayant une structure linéaire, font plutôt appel à une stratégie des connaissances rapportées, moins couteuse d'un point de vue cognitif que la stratégie des connaissances transformées davantage pertinente pour les textes expositifs. Les textes narratifs devraient donc être écrits plus rapidement, c'est-à-dire que leurs pauses initiales et leurs temps d'exécution devraient être plus courts que pour les textes expositifs.

La maitrise des gestes graphomoteurs et des connaissances discursives et linguistiques devraient permettre aux enfants les plus âgés d'écrire plus rapidement. Il est donc attendu que le groupe le plus âgé (3ème) présente des textes avec des pauses et un temps d'exécution plus courts que les deux groupes plus jeunes.

Enfin, les données chronométriques devraient être affectées par les caractéristiques syntaxiques du texte produit. Ainsi, l'emplacement syntaxique devrait affecter la durée de la pause initiale, celle-ci devant être de plus en plus courte devant respectivement une phrase (« P »), une unité Terminale (« T »), une clause (« C ») et un mot interne (« I »). Par ailleurs, si l'on suppose une planification sur plusieurs clauses liées entre elles, on attend que plus une clause dépend de la précédente (HYPO, ENDO), plus la pause la précédant devrait être courte ; au contraire, les clauses matrices devraient être plus longues à planifier. Cependant, il est attendu que cet effet de dépendance syntaxique soit plus important pour les scripteurs les plus âgés, l'hypothèse étant que l'unité de planification devienne plus grande.

1.3. Résultats

Le Tableau 5 indique la durée des pauses avant les mots (en ms) et du temps d'exécution (en ms/car) suivant l'ordre de production, le type de texte et le niveau scolaire, les différences étant illustrées par les Figure 12, Figure 13 et Figure 14. Un modèle linéaire général 2x2x3 a été utilisé afin de voir l'impact de l'ordre de production (écrit1 et écrit2), du type de texte (narratif et expositif) et du niveau scolaire (CM2, 5ème, 3ème) sur les pauses initiales et les temps d'exécution des mots. Des analyses post-hoc ont permis de voir si les effets significatifs de l'ordre de production et du type de texte se retrouvaient pour chaque niveau scolaire.

		Écrit1	Écrit2	Expositif	Narratif	CM2	5ème	3ème
Pause initiale	M	1202,05	1023,90	1238,49	999,48	1244,72	1109,88	984,32
	(SD)	(515,65)	(569,13)	(909,57)	(498,80)	(637,39)	(485,38)	(489,93)
	N	60	60	120	120	40	40	40

Temps d'exécution	M	368,42	371,21	372,87	367,12	452,17	382,09	326,89
	(SD)	(125,48)	(142,78)	(133,96)	(134,57)	(168,15)	(144,48)	(84,04)
	N	315	308	290	333	116	221	286

Tableau 5. Données chronométriques en fonction de l'ordre de production, du type de texte et du niveau scolaire (moyennes en ms, écart-types et effectifs)[46]

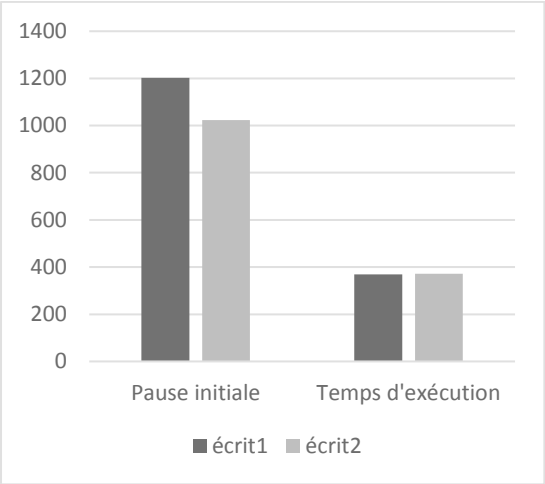


Figure 12. Données chronométriques en fonction de l'ordre de production (moyennes en ms)

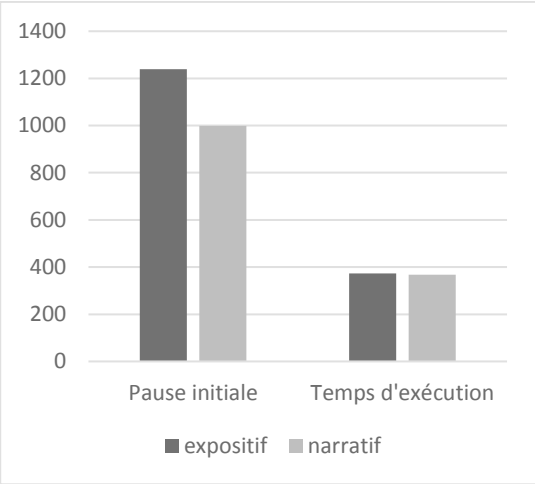


Figure 13. Données chronométriques en fonction du type de texte (moyennes en ms)

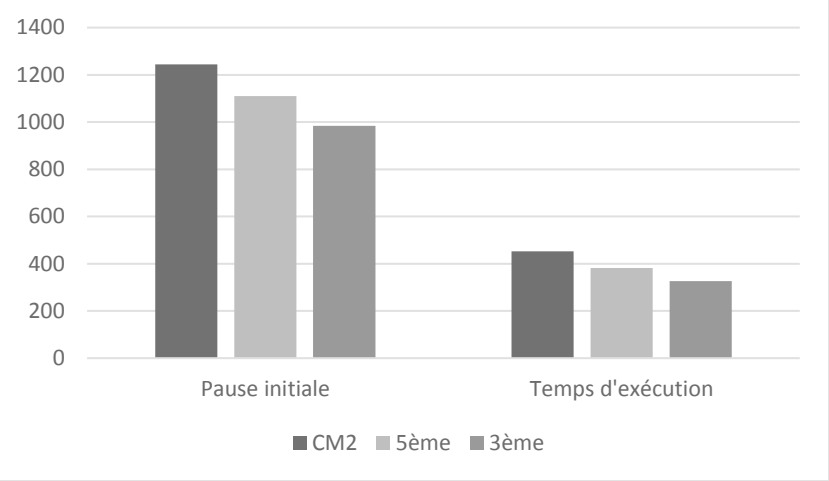
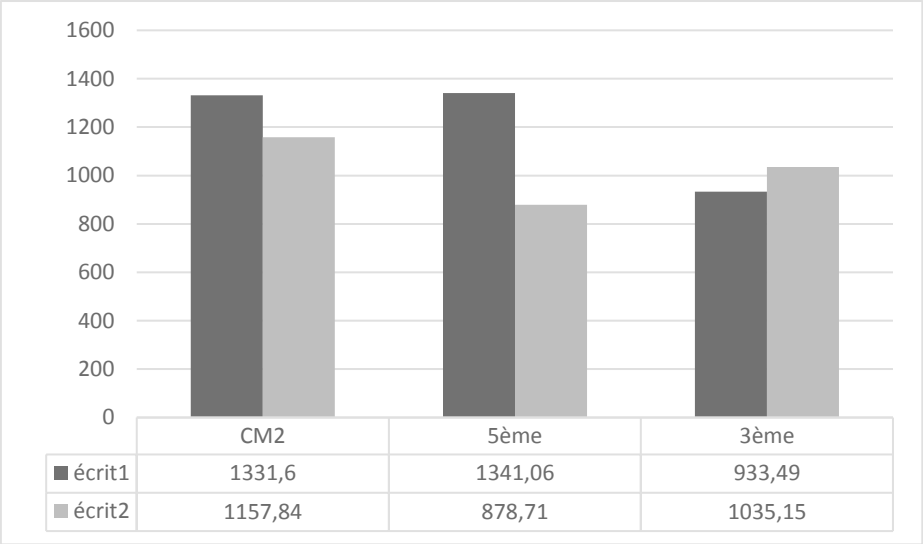


Figure 14. Données chronométriques en fonction du niveau scolaire (moyennes en ms)

1.3.1. Ordre de production

Les pauses avant les mots sont plus courtes lorsque les participants ont produit dans l'ordre oral-écrit ($F_{(1, 114)}$=8,64 ; p=0,004). Il y a une interaction avec le niveau scolaire ($F_{(2, 114)}$=3,417 ; p=0,036), comme cela peut être remarqué sur la Figure 15. Si l'on regarde l'effet de l'ordre de manière indépendante pour chaque niveau scolaire, l'effet est significatif pour les élèves de 5ème ($F_{(1, 35)}$=11,40 ; p=0,002). Aucun effet significatif n'a été relevé concernant le temps d'exécution.

[46] Lorsque nous indiquons « moyennes en ms », il s'agit pour les temps d'exécution de « ms / car ».



	CM2	5ème	3ème
écrit1	1331,6	1341,06	933,49
écrit2	1157,84	878,71	1035,15

Figure 15. Durée des pauses initiales en fonction de l'ordre de production et du niveau scolaire (moyennes en ms)

1.3.2. Type de texte

De manière générale, il y a un effet significatif du type de texte sur les pauses avant les mots ($F_{(1, 228)}$=7,52 ; p= 0,007), qui se retrouve également lorsque l'on ne regarde que les productions des 5ème ($F_{(1, 70)}$=9,22 ; p= 0,004) : ces pauses sont plus courtes dans le cas des textes narratifs. Aucun effet significatif n'apparait pour le temps d'exécution.

1.3.3. Effet du niveau scolaire

Aucun effet général du niveau scolaire n'apparait sur les pauses initiales. Le temps d'exécution du mot (en ms/car) diminue suivant le niveau scolaire ($F_{(2, 610)}$=57,74 ; p<0,001), l'effet étant significatif entre chaque paire (avec toujours p<0,001), ce qui signifie que l'exécution motrice est de plus en plus rapide.

1.3.4. Emplacement syntaxique des pauses

Le Tableau 6 indique la durée moyenne des pauses en fonction de l'emplacement syntaxique. Les t-tests appariés menés pour chaque niveau scolaire en fonction de l'emplacement syntaxique montrent que, quel que soit le niveau scolaire, la pause avant un mot interne à la clause (« I ») est plus courte que celle avant une clause (« C ») et que celle avant une unité T (« T »), comme illustré par la Figure 16. Les t-tests appariés montrent également que, pour les élèves de 5ème et de 3ème, la pause « T » est plus longue si c'est une unité T en début de phrase (« T+P »), qu'une unité T en milieu de phrase (« T-P »). Le nombre insuffisant de phrases graphiques dans les textes des élèves CM2 n'a pas permis l'utilisation de tests statistiques, mais d'après la représentation graphique de la Figure 17, la différence observée pour les élèves de 5ème et de 3ème ne semble pas exister pour ceux de CM2.

	CM2 (N=28)	**5ème** (N=40)	**3ème** (N=40)
« I »[47]	1008,67 (556,78)	759,16 (322,37)	631,24 (244,60)
« C »	2169,74 (1835,21)	1852,80 (1818,00)	1609,26 (1127,62)
« T »	4048,47 (3425,27)	3932,70 (3381,18)	5001,64 (5515,16)
	CM2 (N=4)	**5ème** (N=23)	**3ème** (N=32)
« T-P »	6970,38 (1177,80)	2433,85 (1779,07)	2141,70 (1693,30)
« T+P »	5614,40 (3536,05)	6955,75 (8229,98)	8885,45 (1391,94)

Tableau 6. Durée des pauses initiales en fonction de l'emplacement syntaxique et du niveau scolaire (moyennes en ms, écart-types et effectifs)

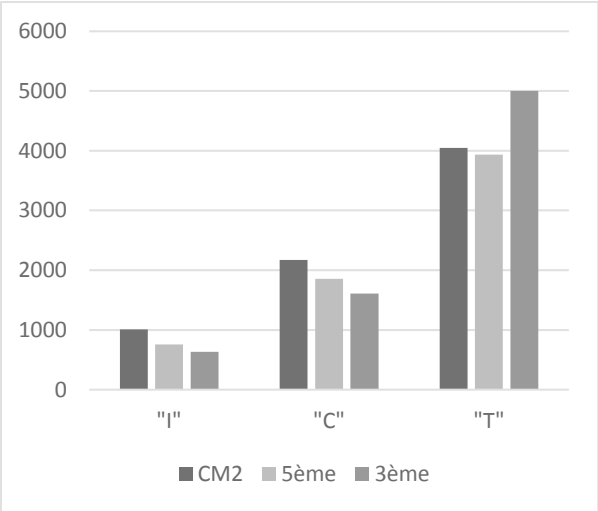


Figure 16. Durée des pauses initiales en fonction des emplacements syntaxiques «I», «C» et «T» et du niveau scolaire (moyennes en ms)

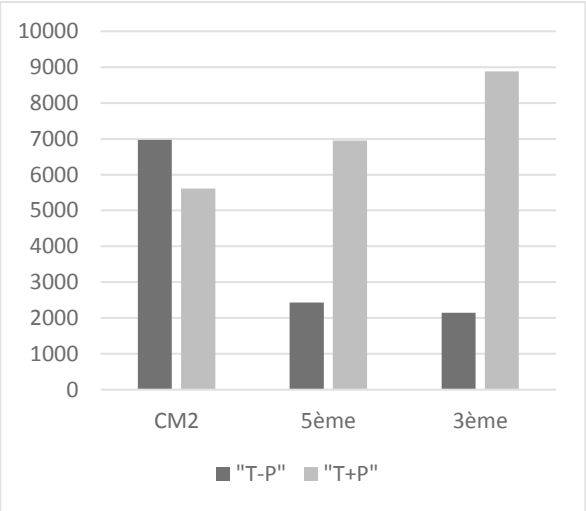


Figure 17. Durée des pauses initiales en fonction des emplacements syntaxiques «T-P» et «T+P» et du niveau scolaire (moyennes en ms)

Par ailleurs, en menant une ANOVA 3x2 pour voir l'effet de l'emplacement syntaxique et du niveau scolaire sur la durée de la pause, on note une interaction entre ces deux facteurs ($F_{(2, 102)}$=4,31 ; p=0,016). Alors que les élèves de 3ème font des pauses plus courtes que ceux de CM2 et de 5ème avant « I » et « C », la durée de la pause augmente devant « T » (Figure 16).

[47] « I » : mot interne à la clause ; « C » : mot en début de clause, mais pas en début d'UT ; « T » : mot en début d'UT ; « T-P » : mot en début d'Unité T mais pas en début de phrase ; « T+P » : mot en début d'Unité T et en début de phrase

1.3.5. Type de clause

		ISO	PARA	HYPO	ENDO	ISO après	HYPO avant
Pause initiale de la clause	M	4227,33	1544,72	1736,26	817,44	1520,26	8031,00
	(SD)	(10284,30)	(3740,80)	(3799,90)	(1224,13)	(2234,38)	(15012,40)
	N	1216	397	409	88	116	49

Tableau 7. Durée des pauses initiales des clauses en fonction du type de clause (moyennes en ms, écart-types et effectifs)

Le Tableau 8 indique la durée des pauses initiales des clauses en fonction du type de clause. Les moyennes ont été calculées à partir de l'ensemble des clauses retenues pour l'analyse, d'où des différences de nombre suivant le type de clause (par exemple, il y a beaucoup plus de clauses PARA que ENDO). Le Tableau 8 présente la durée des pauses initiales en distinguant les clauses indépendantes (ISO) et dépendantes (PARA, HYPO, ENDO) en fonction du fait qu'elles soient matrices ou non matrices. Des ANOVAs à un facteur et des t-tests ont été menés afin de voir si, pour chaque niveau scolaire, il y avait un effet du type de clause sur les pauses initiales. Seuls les résultats significatifs seront donnés ici.

		Indépendante (ISO) matrice	Indépendante (ISO) non-matrice	Dépendante (PARA, HYPO, ENDO) matrice	Dépendante (PARA, HYPO, ENDO) non-matrice
Pause initiale de la clause	M	5281,27	3696,33	2172,89	1419,11
	(SD)	(19688,60)	(8758,66)	(4725,92)	(3283,61)
	N	432	784	168	726

Tableau 8. Durée des pauses initiales des clauses matrices et non-matrices (moyennes en ms, écart-types et effectifs)

Pour tous les niveaux scolaires, des t-tests appariés ont révélé que les clauses indépendantes (ISO) sont précédées d'une pause plus longue que les clauses dépendantes (PARA, HYPO, ENDO) (CM2 : $t_{(16)}$=-3,24 ; p= 0,005 ; 5ème : $t_{(16)}$=-3,24 ; p= 0,005 ; 3ème : $t_{(37)}$=-6,44 ; p< 0,001). La Figure 18 montre que l'écart s'amplifie avec l'âge.

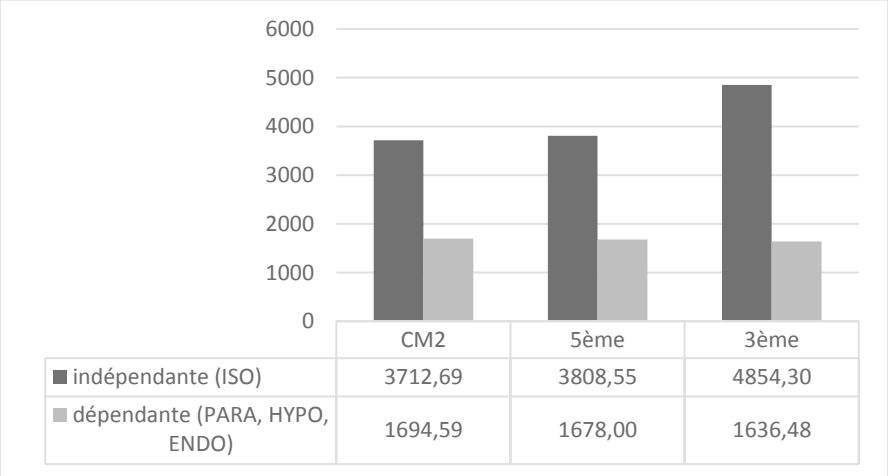
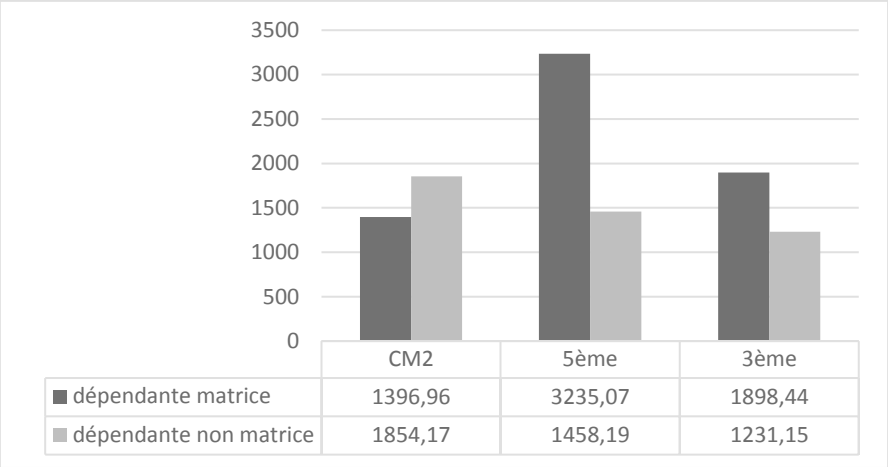


Figure 18. Durée des pauses initiales des clauses indépendantes et dépendantes (moyennes en ms)

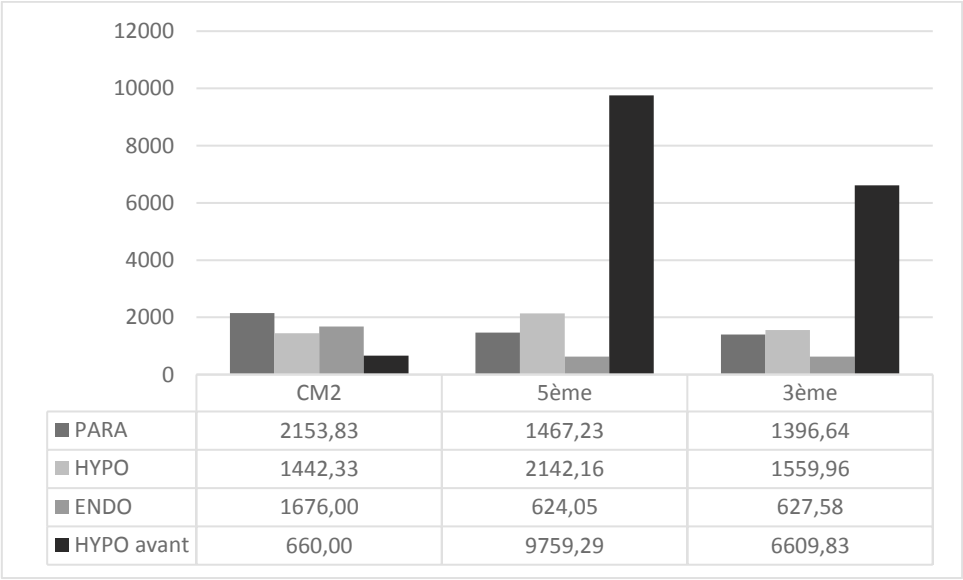
Par ailleurs, les pauses initiales des subordonnées matrices sont plus longues que celles des subordonnées non-matrices de manière significative pour les élèves de 5ème ($t_{(312)}$=2,65 ; p=0,009) et de 3ème ($t_{(135)}$=2,62 ; p=0,01), comme illustré dans la Figure 19.



	CM2	5ème	3ème
dépendante matrice	1396,96	3235,07	1898,44
dépendante non matrice	1854,17	1458,19	1231,15

Figure 19. Durée des pauses initiales des clauses dépendantes matrices et non-matrices (moyennes en ms)

Enfin, une ANOVA à un facteur a permis d'analyser les différences significatives entre les quatre types de clauses dépendantes (PARA, HYPO, ENDO, HYPO_avant). Chez les élèves de 5ème on observe un impact du type de subordonnée sur la durée de la pause initiale : les pauses avant les clauses ENDO sont plus courtes que celles des PARA (p=0,046), des HYPO (p=0,039) et des HYPO_avant (p=0,004). C'est également le cas pour les élèves de 3ème : les pauses initiales des enchâssées sont plus courtes que celles des HYPO (p<0,001) et des HYPO avant (p<0,001). En outre, pour les élèves de 3ème, il apparait que les pauses avant les clauses HYPO_avant sont significativement plus longues qu'avant tous les autres types de clause (pour chaque paire, p<0,01). La Figure 20 illustre ces résultats.



	CM2	5ème	3ème
PARA	2153,83	1467,23	1396,64
HYPO	1442,33	2142,16	1559,96
ENDO	1676,00	624,05	627,58
HYPO avant	660,00	9759,29	6609,83

Figure 20. Durée des pauses initiales des clauses dépendantes en fonction du type de construction syntaxique (moyennes en ms)

1.3.6. Accélération de l'écriture au niveau de la clause et de l'unité T

Des analyses de régressions ont été menées afin de voir si un schéma d'accélération ou de ralentissement apparaissait lors de l'écriture de la clause et de l'unité T. Dans cette analyse, pour chaque niveau scolaire, la VD était la durée de la pause avant le mot, et les VI sa position dans la clause et dans l'unité T. Beaucoup de facteurs ayant un impact sur cette durée de la pause, le pourcentage de variance expliquée par la position du mot dans la clause est assez faible (pseudo R^2=4,2% pour les CM2, 4,4% pour les 5ème et 3,4% pour les 3ème). Cependant, pour les trois niveaux scolaires l'effet est significatif ($p<0,001$), et le coefficient β est négatif, ce qui veut dire que le temps de pause diminue lorsque l'on se rapproche de la fin de la clause. Pour la position du mot dans l'unité T, l'effet n'est significatif que pour les élèves de 3ème ($p<0,001$, la variance expliquée en plus est pseudo R^2=0,001%, β est négatif) : ces derniers montrent donc également un effet d'accélération en fin d'unité T.

1.4. Discussion

La question principale guidant cette étude était de savoir dans quelle mesure l'unité de production varie selon l'âge, en prenant en compte l'influence potentielle du type de texte et de l'ordre de production. Pour cela, nous avons utilisé les données chronométriques telles que la durée des pauses initiales et le temps d'exécution du mot. En effet, lorsque trop de processus cognitifs sont engagés, le scripteur peut éprouver des difficultés à continuer à écrire, et son geste graphique va ralentir, voire s'arrêter.

Concernant l'ordre de production, l'hypothèse générale selon laquelle les pauses initiales seraient plus courtes lorsque le scripteur effectue sa production écrite après une production orale se trouve confirmée. Cela témoigne d'un effet facilitateur de la première production : le participant ayant déjà réfléchi au contenu de son texte, il effectue des pauses moins longues. Cependant, dans le détail, on remarque que cet effet ne se trouve que chez les enfants de 5ème. L'absence d'effet pour les élèves de CM2 peut s'expliquer par le fait que leurs ressources cognitives sont fortement mobilisées par le geste graphique et les contraintes orthographiques, ces élèves ne tirent donc pas de bénéfice d'une planification préalable. Les élèves de 3ème, quant à eux, sont davantage familiers de la modalité écrite. Ils en connaissent donc les spécificités, et leur langage écrit diffère de celui oral (Berman & Ravid, 2009). Leur attention est alors essentiellement focalisée sur la sélection de structures et termes adéquats à la modalité écrite. L'exemple (47) montre la version écrite et orale d'un texte d'un élève de 3ème dans laquelle la même information est communiquée, mais avec des structures syntaxiques très différentes.

(47) *Mon frère m'a conseillé d'acheter le portable que vendait la sœur de son copain* [3ème,1,f,n,écrit2 - version écrite]

(48) c'était pour une histoire de portable je me rappelle c'était euh mon frère en fait il m'avait trouvé un portable c*'était la c'était la euh la petite sœur de son copain* ben elle vendait son portable [3ème,1,f,n,écrit2 - version orale]

Les élèves de 5ème, pour qui la sélection orthographique et l'exécution motrice sont relativement automatisés, n'ont probablement pas encore établi les particularité du langage oral et écrit (Mazur-Palandre, 2015). Ainsi, ils peuvent tirer un avantage du fait d'avoir planifié le contenu au préalable, lors de la production orale.

Le même pattern de résultats est visible pour l'impact du type de texte. Nous faisions l'hypothèse que l'écriture d'un texte narratif, comparé à un texte expositif, s'effectuerait avec des pauses plus courtes et une exécution plus rapide. De nouveau, seules les productions des 5ème ont des pauses plus courtes pour les textes narratifs qu'expositifs. Van Hell et al. (2008) avaient déjà remarqué qu'il n'y avait pas de différence suivant le type de texte pour les élèves de 6ème, alors que pour les adultes les pauses du texte expositif tendaient à être plus longues que celles du texte narratif. Il est alors possible que les élèves de CM2, comme les néerlandais de 6ème, soient encore trop contraints par le geste graphique pour que les pauses soient influencées par une facilitation des processus de haut-niveau. La durée des pauses dans les productions des élèves de 3ème ne différait pas non plus suivant le type de texte. Or, leurs textes sont relativement hybrides, comme l'avait remarqué (Berman, 2004a) : par exemple, il peut leur arriver d'illustrer leurs discours expositifs par des exemples personnels et d'exprimer des informations génériques ou impersonnelles dans leurs textes narratifs. Cela pourrait expliquer l'absence de variation de la durée des pauses en fonction du type de texte. De nouveau, seuls les élèves de 5ème semblent tirer bénéfice de l'effet facilitateur du texte narratif.

Contrairement à notre hypothèse, la durée des pauses initiales ne diminue pas avec l'âge. Nous attendions une telle diminution du fait que, grâce à l'automatisation de nombreux processus rédactionnels, la diminution de leur cout cognitif se reflèterait par des pauses moins longues. En réalité, l'automatisation des processus de bas-niveau est bien visible à travers l'accélération de l'exécution motrice, puisque le temps mis pour écrire un mot est de plus en plus court. D'ailleurs, lorsque l'on regarde la durée des pauses initiales en fonction de l'emplacement syntaxique, on voit que les pauses précédant les mots internes aux unités T diminuent : seule la pause en début d'unité T augmente fortement chez les élèves de 3ème. Or, il est difficile de déterminer exactement ce qui se passe durant cette pause, mais il est probable que son rôle change avec l'âge. De plus, il est possible que la quantité d'informations liées au thème de l'écriture soit plus importante, rendant leur sélection plus longue. Ces pauses initiales d'unité T semblent alors être bien le reflet de processus de plus haut niveau, comme la planification, ces processus restant couteux cognitivement même pour les élèves plus âgés. Cela rejoint

le constat de certaines études précédentes (McCutchen, 2006; Scardamalia & Bereiter, 1991) soulignant que l'activité rédactionnelle reste difficile malgré l'automatisation de certains processus, car les scripteurs mettent en place des stratégies permettant de produire des textes de meilleure qualité.

Parmi les résultats obtenus, plusieurs suggèrent que l'unité de production augmente avec l'âge : alors que les plus jeunes planifient essentiellement clause par clause, les plus âgés planifient plutôt au niveau de l'unité T, voire de la phrase.

Les études mettant en relation la structure hiérarchique du texte et la durée des pauses ont bien mis en évidence que la clause était une unité importante en production (entre autres Chanquoy et al., 1990; Foulin, 1998a; van Hell et al., 2008). Dans notre étude, nous voyons également que tous les scripteurs font des pauses plus longues avant l'unité T qu'avant la clause ; cependant, cette pause est beaucoup plus longue pour les élèves de 3ème que pour ceux de CM2 et de 5ème, alors que l'ensemble des autres pauses diminue. En outre, les pauses sont beaucoup plus longues en début de phrase qu'en début de d'unité T interne à la phrase pour les élèves de 5ème et de 3ème. Pour les élèves de CM2, en revanche, l'unité T et la phrase ne semblent pas être des unités importantes de planification, ce qui va dans le sens de ce qu'avait remarqué Foulin (1998) pour des enfants plus jeunes.

Par ailleurs, si la planification se fait sur plusieurs clauses, la durée des pauses avant les clauses devrait varier suivant le type de clause (Immonen & Mäkisalo, 2010; van Hell et al., 2008). En effet, si un scripteur sait par exemple que sa clause va en contenir une enchâssée et qu'il planifie au-delà d'une clause, la pause avant la clause matrice sera plus longue, alors que celle avant la clause enchâssée sera relativement courte, puisque celle-ci aura déjà été planifiée. Tous les élèves effectuent des pauses plus courtes devant les clauses dépendantes (PARA, HYPO, ENDO) qu'indépendantes, mais la différence entre ces deux types de clause s'amplifie pour les élèves les plus âgés. De même, les pauses devant des subordonnées matrices sont plus longues que devant les subordonnées non-matrices pour les élèves de 5ème et de 3ème : durant cette longue pause, il est probable que le scripteur débute la planification de la clause subordonnée qui suivra la clause matrice. Enfin, les clauses ayant déjà été planifiées, comme les enchâssées (ENDO), sont précédées d'une pause plus courte pour les élèves les plus âgés. Les élèves de 3ème, quant à eux, effectuent les pauses les plus longues devant les subordonnées placées en début d'unité T, cette pause initiale permettant à la fois la planification de la subordonnée, mais également de la clause matrice qui ici est située après la subordonnée. Il apparait donc que, dès la 5ème et encore plus pour les élèves de 3ème, les scripteurs cessent de planifier uniquement clause par clause, pour planifier souvent ensemble les clauses qui sont dépendantes syntaxiquement. Cette capacité à planifier de manière plus large est sans doute liée à l'automatisation de certains processus de bas-niveaux (Berninger & Swanson, 1994; Olive, Favart, Beauvais & Beauvais, 2009), ainsi qu'à l'augmentation de la capacité de la MDT, qui permet aux scripteurs plus âgés d'augmenter leur empan de planification.

Enfin, si l'unité de production augmente effectivement avec l'âge, on devrait en trouver des traces dans le rythme d'écriture global de la clause et de l'unité T. En effet, en début d'unité les ressources cognitives sont fortement mobilisées pour la planification et pour maintenir en mémoire l'ensemble de l'unité planifiée. La libération de la mémoire au fur et à mesure de l'écriture de l'unité peut permettre une exécution motrice plus aisée, donc plus rapide. Or, les analyses de régression ont effectivement montré une corrélation négative entre la position du mot dans la clause et la pause qui le précède : les pauses sont ainsi de plus en plus courtes, ce qui signifie que, quel que soit le niveau scolaire, il y a un schéma d'accélération en fin de clause. Or, on constate que, pour les élèves de 3ème, un tel schéma d'accélération apparait également en fin d'unité T : il semble donc bien que, pour ces scripteurs, une planification globale existe pour cette unité.

Ces résultats nous renseignent donc sur le développement des stratégies rédactionnelles, grâce à l'analyse en temps réel des productions écrites. Pour les élèves de CM2, dont de nombreuses ressources cognitives sont mobilisées pour des processus de bas-niveau encore non automatisés, la modalité écrite est difficile. Ces contraintes de bas-niveau les obligent à réfléchir au texte clause par clause, conduisant majoritairement à des constructions juxtaposées ou coordonnées. Les élèves de 5ème, pour qui le geste graphomoteur est moins difficile, parviennent à tirer profit d'une première production orale ; de manière générale leur rythme d'écriture est plus rapide que pour les élèves de CM2. On voit également que cela leur permet, peu à peu, de passer d'une planification portant uniquement sur la clause, à une planification plus large s'étendant à l'unité T. Enfin, la familiarité avec la modalité écrite permet aux élèves de 3ème d'adopter des nouvelles stratégies rédactionnelles. Ces derniers privilégient en effet des pauses plus longues en début de phrase, planifiant leur production sur plusieurs clauses lorsque celles-ci entretiennent des liens de dépendance syntaxique. Ainsi, les analyses issues des données chronométriques montrent l'augmentation de l'unité de planification avec l'âge.

2. La phrase : de la maitrise de ses conventions graphiques à sa pertinence cognitive[48]

2.1. Introduction

Lors de l'étude précédente, qui avait pour objectif d'identifier quelles étaient les unités de planification, la phrase s'est révélée une unité pertinente pour expliquer la dynamique de l'écriture pour les scripteurs les plus avancés. En effet, l'analyse des données chronométriques (pauses et temps d'exécutions) de ces rédactions a montré qu'au fur et à mesure qu'ils avancent dans leur scolarité, les scripteurs semblent planifier leur discours non plus clause par clause, mais à un niveau supérieur, comme la phrase syntaxique[49]. Cette dernière semble alors devenir une unité réellement pertinente du point de vue cognitif lors de la planification du texte, ce qui rapproche le comportement des élèves de 3ème de celui observé chez des adultes. La manière dont le scripteur segmente son texte afin de le planifier est vraisemblablement marquée par la ponctuation utilisée, puisque celle-ci, d'un point de vue cognitif, « constitue le lieu privilégié d'accomplissement de l'activité de planification » (Passerault, 1991, p. 25), même si elle est souvent vue comme une aide pour le lecteur, et est donc parfois rajoutée a posteriori lors de la relecture du texte. Or, la façon dont les clauses sont assemblées ou séparées pour former des unités (par l'usage de connecteurs et de la ponctuation) ne semble pas chose aisée du côté des scripteurs débutants aussi bien que du côté de l'enseignement. Un aspect délicat de la didactique de l'écrit concerne notamment l'enseignement de la ponctuation, dont la plurifonctionnalité est parfois peu identifiée par les enseignants (Dufour, 2014), et qui est en outre étroitement liée à la compréhension de la notion de « phrase », elle-même difficile à définir (Béguelin, 2002; Berrendonner, 2002; Kleiber, 2003).

Les frontières d'unités syntaxiques méritent donc d'une part d'être analysées par rapport aux normes de marquage typographique, et d'autre part d'être abordées du point de vue de leur fonction cognitive. La question se pose ainsi du rapport entre la maitrise de l'ensemble des conventions liées à la phrase graphique canonique et la pertinence de celle-ci d'un point de vue cognitif. Nous nous proposons donc de mettre en lien les données chronométriques et le rapport à la norme que présentent les marques de ponctuation.

---

[48] Cette étude reprend un article à paraitre dans la revue LIDIL en décembre 2016, dans le n°54 dont la thématique est « La phrase en production d'écrits, approches nouvelles en didactique ».

[49] Cette étude devant paraitre dans une revue dont la thématique était la « phrase », nous avons préféré utiliser les termes « phrase syntaxique » et « phrase graphique », plus explicites dans un cadre de réflexion didactique. Ces termes recouvrent respectivement les définitions de « unité T » et « phrase », selon la terminologie utilisée dans cette thèse.

2.2. Méthodologie

2.2.1. Données utilisées

Les textes étudiés sont issus du corpus Reflex. Les scripteurs, âges de 10 à 15 ans, étaient répartis en trois groupes de quarante participants, suivant le niveau scolaire (CM2, 5ème, 3ème.). Deux types de textes étaient produits – un expositif et un narratif – sur le thème des problèmes entre les gens. Bien que chaque texte ait été produit à l'écrit et à l'oral, seuls les textes écrits sont utilisés pour cette étude ; notre corpus est donc constitué de 240 textes (3 niveaux * 40 participants * (1 texte expositif + 1 texte narratif)). Ces textes étaient rédigés sur une tablette à digitaliser et enregistrés grâce au logiciel Eye & Pen.

2.2.2. Codage des données

Trois types d'informations ont été codés : la segmentation du texte, les écarts par rapport à la norme concernant la ponctuation et la longueur des pauses aux frontières d'unités syntaxiques.

2.2.2.1. Segmentation du texte

Chaque texte a été découpé selon plusieurs unités, telles que définies précédemment : les phrases graphiques (PG), débutant par une majuscule et se terminant par un point[50] ; les phrases syntaxiques (PS), qui sont des paquets de clauses en relation de dépendance syntaxique[51]; les clauses (C), au sens de Berman et Slobin (1994), unités comportant un seul prédicat unifié.

L'énoncé (49) correspond à une phrase graphique (PG), composée elle-même de deux phrases syntaxiques (PS), séparées par « // » ; la deuxième PS est composée de deux clauses (C), séparées par « / ».

(49) À deux heures trente tout le monde repartit en cours // et on apprit tous / que la principale appellerait nos parents[52]. [S3, 3ème, narratif][53]

Pour la suite de notre étude, nous utiliserons les termes PG, PS et C de manière mutuellement exclusive : les « frontières de C » correspondent à un début de nouvelle clause qui ne soit pas une nouvelle PS ; les « frontières de PS » correspondent à un début de nouvelle PS qui ne soit pas une nouvelle PG. Cela est illustré dans l'exemple (50), qui reprend l'énoncé de l'exemple (49).

---

[50] Dans notre corpus, il arrivait qu'une des deux marques soit absente : une nouvelle PG était tout de même comptée dès lors qu'il y avait un point ou une majuscule.

[51] Ces phrases syntaxiques sont définies exactement de la même manière que les unités T.

[52] Les erreurs d'orthographe n'ont pas été conservées.

[53] Les exemples sont identifiés avec les éléments suivants : [n° du sujet, niveau scolaire, type de texte]

(50) [PG] À deux heures trente tout le monde repartit en cours [PS] et on apprit tous [C] que la principale appellerait nos parents. [S3,3ème,narratif]

Pour chaque texte, nous avons relevé la proportion de PG contenant 1 ou 2 PS : cela correspond à des PG plutôt « canoniques », comme ce que l'on pourrait trouver en majorité dans des textes d'adultes.

2.2.2.2. Usage des marques graphiques de segmentation

Il arrive fréquemment que les participants utilisent la ponctuation de façon non standard lorsqu'ils segmentent leur texte. Dans ce cadre, nous avons identifié deux types de problèmes. Premièrement, il peut y avoir un problème au niveau du marquage des PG, avec soit une absence de majuscule (51) soit une absence de point (52). En effet, la présence d'une seule des deux marques était suffisante pour identifier une PG, d'où la possible absence de la seconde marque.

(51) *Je me disais que mon père m'aimerait encore* moins. ça *m'a fait un sacré choc.* [S34, CM2, narratif]

(52) *J'étais assis sur un banc alors que je parlais avec mon copain, une série de coups* de pied me frappa Tous étaient mes amis sauf un. [S46, 5ème, narratif]

Deuxièmement, nous avons identifié les problèmes de marquage entre 2 PS : celles-ci ont été segmentées en fonction de critères syntaxiques, mais il arrivait qu'elles ne soient séparées ni par une virgule, ni par un connecteur, ce qui n'est pas canonique (53).

(53) Il commençait à faire *nuit c'*était bientôt l*'heure de manger.* [S34, 3ème, narratif]

2.2.2.3. Longueur des pauses

En regard du codage des marques de segmentation ou de leur absence, nous avons examiné la longueur des pauses associées. Pour cela, nous avons pris en compte la pause entre le dernier mot de l'unité et le premier mot de l'unité suivante. Par exemple, dans le cas de deux PG séparées par un point, la « pause entre PG » est constituée de la pause avant le point additionnée de la pause après le point ; si le point était omis entre les deux PG, il s'agissait simplement de la pause entre les deux mots.

Dans un premier temps, nous avons identifié les pauses les plus longues pour chaque texte. Plutôt que de déterminer un seuil valable pour tous les sujets, nous avons déterminé des seuils pour chaque texte. Pour chaque texte, nous avons classé toutes les pauses (c'est-à-dire le temps écoulé entre deux mots) dans l'ordre décroissant. Le seuil pour les pauses les plus longues que nous avons retenu correspond à la durée au-delà de laquelle il y a les 20% de pauses les plus longues. Ce seuil nous a permis de distinguer les pauses « longues » du texte et les pauses « courtes » : une pause est « longue » si elle fait partie des 20% de pauses les plus longues du texte ; elle est « courte » si elle fait partie des 80% de pauses les moins longues.

Dans un second temps, nous avons déterminé, pour chaque emplacement syntaxique, dans quelle catégorie de pause se situe le mot : par exemple, parmi toutes les pauses en frontière de PG dans un texte, quel est le pourcentage de pauses qui sont longues. Ainsi, nous avons trois variables exprimées en pourcentages : « frontières de PG – longues », « frontières de PS – longues », « frontières de C – longues ».

2.2.2.4. Exemple de texte codé

Le texte présenté dans l'exemple (54) permet d'illustrer les trois types d'informations codés.

(54) **[PG]** Les gens ne doivent pas faire la guerre **[PS]** car ils se tuent entre eux. **[PG]** Dans les collèges, lycées, écoles il ne faut pas [C] copier sur le voisin [PS] car après *on a des problèmes dans l'école* **[PS]** *on a copié sur le voisin [PS] donc on n'apprend* rien. **[PG]** Les élèves pourraient éviter de se bagarrer pour des histoires **[PS]** car quand on est grand après [C] on se tue avec des armes.

Légende :

[PG] : frontière de phrase graphique ; [PS] : frontière de phrase syntaxique ; [C] : frontière de clause

frontière en gras : pause longue (parmi les 20% des pauses les plus longues du texte)

frontière soulignée : problème de ponctuation

2.2.3. Analyses utilisées et présentation des résultats

Nous avons tout d'abord effectué des ANOVAs, avec pour variable indépendante le niveau scolaire (3 modalités : CM2, 5ème, 3ème). Dans le cas d'un effet significatif du niveau scolaire, des analyses post-hoc ont permis de déterminer si la différence était significative entre chaque niveau scolaire.

Nous avons ensuite établi des « profils », grâce aux analyses « en nuées dynamiques » (ou k-means clusters), en utilisant le logiciel SPSS. Pour cela, nous avons à chaque fois fait des essais avec 2, 3 et 4 groupes, puis nous avons conservé le nombre de groupes qui paraissait le plus pertinent, au vu du nombre d'individus par groupe et des caractéristiques des profils ainsi déterminés. Les graphiques proposés ont été effectués à partir des centres de classe finaux.

2.3. Résultats

2.3.1. Caractéristiques générales

2.3.1.1. La segmentation

Nous avons comptabilisé le nombre moyen de mots, de phrases syntaxiques et de phrases graphiques par texte. Le Tableau 9 indique les moyennes et les écart-types de chaque variable en fonction du niveau scolaire.

Il existe un effet du niveau scolaire pour le nombre de mots par texte ($F_{(2)}$=30,60 ; p<0,001). La différence est significative entre chaque niveau scolaire : les élèves de 3ème écrivent plus de mots que ceux de 5ème, qui eux-mêmes en écrivent davantage que ceux de CM2. Concernant le nombre de PG par texte, un effet du niveau scolaire ($F_{(2)}$=28,74 ; p<0,001) révèle que les textes des 3ème contiennent plus de PG que ceux des 5ème, qui eux-mêmes en contiennent plus que ceux des CM2. Un effet significatif du niveau scolaire indique que les textes des élèves de 5ème et 3ème sont composés de plus de PS que ceux des élèves de CM2 ($F_{(2)}$=21,60 ; p<0,001). Enfin, le pourcentage de PG contenant une ou deux PS dépend aussi du niveau scolaire ($F_{(2)}$=10,36 ; p<0,001) : ce pourcentage est plus élevé dans les textes de 5ème et 3ème que dans ceux de CM2.

	CM2	5ème	3ème
nombre de mots par texte	64,45 (34,39)	124,48 (60,13)	163,60 (70,57)
nombre de phrases graphiques (PG) par texte	2,25 (1,36)	5,13 (3,21)	6,65 (2,95)
nombre de phrases syntaxiques (PS) par texte	5,60 (2,98)	10,63 (6,04)	12,10 (4,38)
% de PG contenant 1 ou 2 PS	54,74% (32,54%)	72,70% (18,77%)	77,61% (16,38%)

Tableau 9. Segmentation des textes en fonction du niveau scolaire (moyennes et écart-types)

2.3.1.2. Seuil des pauses longues

Pour chaque texte un seuil identifiant les 20% de pauses les plus longues a été déterminé. Le Tableau 10 en indique les moyennes et écart-types en fonction du niveau scolaire.

Un effet du niveau scolaire a été relevé ($F_{(2)}$=7,60 ; p<0,001) : le seuil déterminant les « pauses longues » est significativement plus élevé dans les textes des élèves de CM2 que de 5ème et 3ème.

	CM2	5ème	3ème
seuil pauses longues (en ms)	1763,23 (905,55)	1301,43 (671,182)	1136,55 (629,68)

Tableau 10. Seuil des 20% des pauses les plus longues des textes en fonction du niveau scolaire (moyennes et écart-types)

2.3.2. Frontières syntaxiques

2.3.2.1. Usage des marques graphiques de segmentation

Le Tableau 11 indique les moyennes et écart-types des trois types de problèmes liés à la ponctuation : absence de virgule et de connecteur à une frontière de PS, absence de majuscule en début de PG ou de point en fin de PG.

Les problèmes de ponctuation en frontière de PS dépendent du niveau scolaire ($F_{(2)}$=9,06 ; p<0,001) : il y a une différence significative du pourcentage d'erreurs entre les textes des CM2 et ceux des 5ème et 3ème, les plus jeunes faisant davantage d'erreurs. Concernant les problèmes de majuscule, il y a un effet du niveau scolaire ($F_{(2)}$=6,65 ; p<0,005) : les élèves de CM2 font davantage d'erreurs que ceux de 5ème et 3ème. Enfin, il y a également un effet du niveau scolaire pour l'absence de point en fin de PG ($F_{(2)}$=7,94 ; p<0,001), les élèves de CM2 faisant davantage ce type d'écart que ceux de 5ème et 3ème.

	CM2	5ème	3ème
PS juxtaposées sans virgule ni connecteur (% par texte)	24,68% (21,75%)	14,87% (19,54%)	7,04% (13,44%)
PG sans majuscule au début (% par texte)	21,46% (26,48%)	10,22% (15,60%)	5,86% (15,00%)
PG sans point à la fin (% par texte)	12,47% (18,97%)	1,01% (4,28%)	5,05% (11,50%)

Tableau 11. Proportion de problèmes liés à la ponctuation en fonction du niveau scolaire (moyennes en % et écart-types)

Nous avons voulu déterminer des profils de scripteurs en fonction des problèmes de ponctuation. Les analyses menées ont permis d'identifier 3 profils (A1, A2, A3). La Figure 21 illustre les caractéristiques des problèmes de ponctuation pour chaque profil établi.

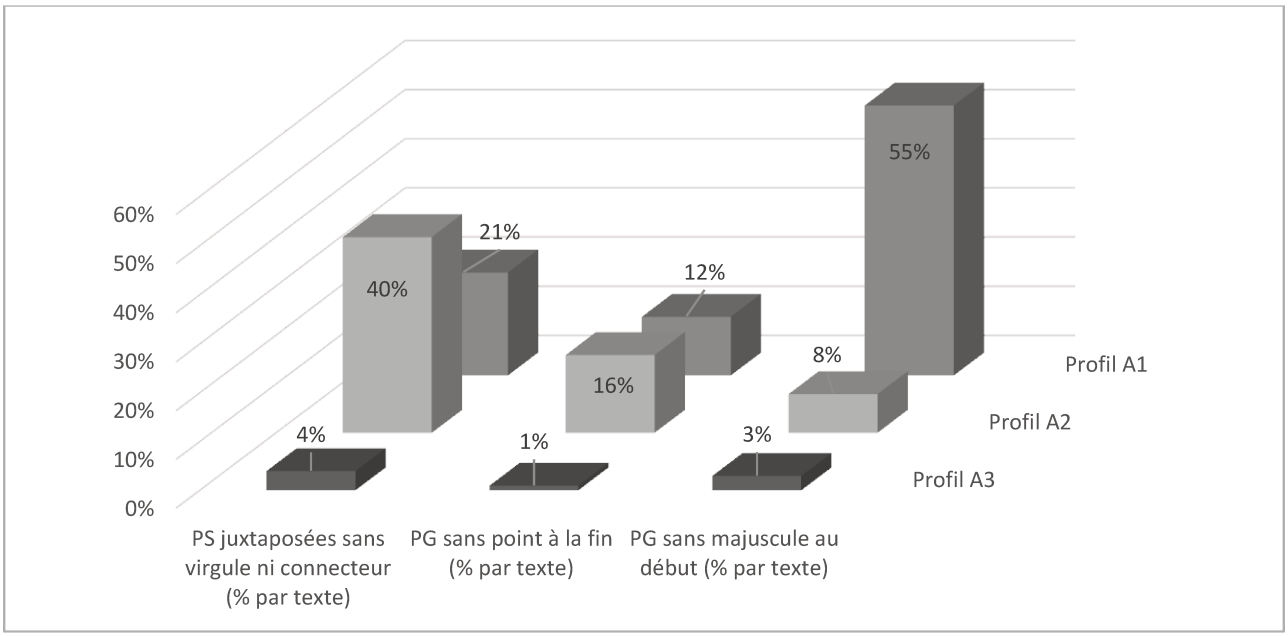


Figure 21. Profils liés aux problèmes de ponctuation

Le profil A3 correspond à l'écriture la plus canonique, puisqu'il y a peu de problèmes de ponctuation. Le profil A2 correspond aux élèves faisant essentiellement des erreurs de virgules (absence de virgules et de connecteurs en frontière de PS). Enfin, les élèves du profil A1 sont ceux qui s'écartent le plus de la norme d'un point de vue quantitatif, leurs difficultés principales se situant au niveau de l'emploi des majuscules en début de phrase graphique.

Le Tableau 12 indique le nombre de scripteurs par profil en fonction du niveau scolaire : par exemple, très peu d'élèves de 3ème ont le profil A1 ou A2, ces deux profils étant surtout répandus parmi les élèves de CM2.

	A1 : beaucoup de problèmes de majuscules	A2 : beaucoup de problèmes de virgules	A3 : peu de problèmes de ponctuation
CM2	13	17	10
5ème	4	8	28
3ème	2	4	34
Total	19	29	72

Tableau 12. Répartition des scripteurs dans les profils liés aux problèmes de ponctuation en fonction du niveau scolaire

2.3.2.2. Longueur des pauses suivant l'emplacement syntaxique

Pour chaque emplacement syntaxique (frontière de PG, de PS ou de C) nous avons relevé la proportion de pauses longues. Les moyennes et écart-types sont donnés dans le Tableau 13.

Il y a un effet du niveau scolaire ($F_{(2)}$=40,64 ; p<0,001) pour la proportion de frontières de PG – longues, avec une augmentation de cette proportion en fonction du niveau scolaire, la différence étant significative entre chaque groupe. En revanche, il n'y a aucun effet significatif pour la proportion de frontières de PS – longues. Concernant la proportion de frontières de C – longues, un effet du niveau scolaire a été relevé ($F_{(2)}$=2,94 ; p=0,05) : cette proportion est plus élevée chez les élèves de CM2 que chez les élèves de 5ème et 3ème.

	CM2	5ème	3ème
Frontière de PG – longues	40,36% (26,70%)	65,91% (17,38%)	78,13% (9,02%)
Frontière de PS – longues	56,16% (28,92%)	56,26% (20,28%)	58,17% (20,11%)
Frontière de C – longues	41,85% (24,43%)	32,25% (15,42%)	35,50% (11,79%)

Tableau 13. Proportions de la longueur des pauses suivant l'emplacement syntaxique en fonction du niveau scolaire (moyennes en % et écart-types)

Nous avons de nouveau établi des profils de scripteurs, cette fois en fonction de la longueur des pauses aux différents emplacements syntaxiques. La Figure 22 illustre les caractéristiques des longueurs des pauses des scripteurs des trois profils obtenus (B1, B2, B3).

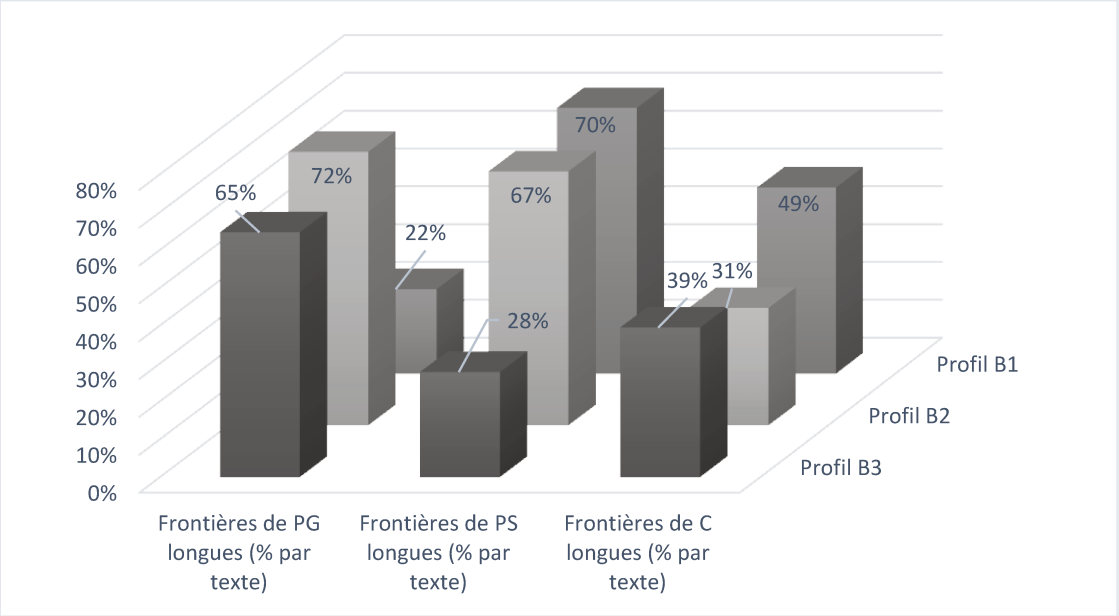


Figure 22. Profils liés à la longueur des pauses suivant l'emplacement syntaxique

Le profil B3 correspond à des scripteurs marquant beaucoup de pauses longues entre les PG (65%), et davantage de pauses courtes entre les PS et les clauses. Ce profil est l'inverse du profil B1, dans lequel les scripteurs marquent peu de pauses longues entre les PG (22%), mais beaucoup de pauses longues entre les PS et un nombre assez élevé de pauses longues entre clauses. Les scripteurs du profil B2, quant à eux, font beaucoup de pauses longues entre les PG et les PS (respectivement 72% et 67%), et plus de pauses courtes entre les clauses.

Le Tableau 14 indique le nombre de scripteurs par profil en fonction du niveau scolaire : par exemple, aucun élève de 3ème n'a le profil B1, ce profil étant surtout trouvé chez les élèves de CM2.

	B1: frontières de PG courtes, frontières de PS longues	B2 : frontières de PG et de PS longues	B3: frontières de PG longues
CM2	17	10	13
5ème	4	27	9
3ème	0	29	11
Total	21	66	33

Tableau 14. Répartition des scripteurs dans les profils liés à la longueur des pauses suivant l'emplacement syntaxique en fonction du niveau scolaire

2.3.3. Lien entre produit final et gestion chronométrique

Nous avons enfin souhaité voir si, de manière générale, les scripteurs ayant des problèmes d'utilisation de la ponctuation étaient ceux qui avaient une gestion des pauses moins canonique (avec peu de pauses longues entre PG). Nous avons donc établi le Tableau 15, qui croise les deux profils, en indiquant le nombre d'élèves par niveau scolaire pour chaque case. Par exemple, 21 élèves ont peu de

problèmes de ponctuation (profil A3) et des frontières de PG longues (profil B3), dont 5 sont en CM2, 6 sont en 5ème et 10 sont en 3ème.

	A1: beaucoup de problèmes de majuscules	A2: beaucoup de problèmes de virgules	A3: peu de problèmes de ponctuation	Total
B1: frontières de PG courtes, frontières de PS longues	7 (7 en CM2)	8 (7 en CM2, 1 en 5ème)	6 (3 en CM2, 3 en 5ème)	21
B2: frontières de PG et PS longues	8 (3 en CM2, 3 en 5ème, 2 en 3ème)	13 (5 en CM2, 5 en 5ème, 3 en 3ème)	45 (2 en CM2, 19 en 5ème, 21 en 3ème)	66
B3: frontières de PG longues	4 (3 en CM2, 1 en 3ème)	8 (5 en CM2, 2 en 5ème, 1 en 3ème)	21 (5 en CM2, 6 en 5ème, 10 en 3ème)	33
Total	19	29	72	120

Tableau 15. Croisement des profils liés à la ponctuation et à la longueur des pauses

2.4. Discussion

Nous nous sommes attachés ici à la segmentation de textes d'élèves telle qu'elle se manifeste à travers la ponctuation et l'organisation syntaxique des clauses en tentant de la mettre en lien avec des mesures chronométriques de l'activité d'écriture. Dans ce travail, nous avons d'abord établi que les textes s'allongent avec les années de scolarisation (niveau scolaire). Les textes des élèves de CM2 sont en effet en moyenne deux fois moins longs (en nombre de mots) que les textes des élèves de 5ème et les textes des élèves de 3ème sont en moyenne 30% plus longs que ceux de 5ème. Si les textes sont plus longs, cela implique plus de possibilités de segmentation et l'on n'est pas étonné de constater des rapports similaires entre CM2, 5ème et 3ème en ce qui concerne le nombre de PG : les élèves de CM2 ont en moyenne 2 fois moins de PG que ceux de 5ème et les élèves de 3ème ont environ 30% de PG de plus que ceux de 5ème. Les proportions ne sont pas si bien conservées pour les PS puisque si les CM2 ont bien environ deux fois moins de PS par texte que les 5ème, les textes des élèves de 3ème comportent en moyenne seulement environ 14% de plus de PS que les textes de 5ème. Ces chiffres suggèrent que la taille des unités de segmentation varie peu mais il conviendrait d'examiner plus particulièrement la longueur des PG et des PS à travers les textes. Enfin, la proportion de PG contenant une ou deux PS, que nous avons identifiée comme un indice de conformité à la norme de segmentation, augmente selon le niveau scolaire, surtout entre les élèves de CM2 et ceux de 5ème et 3ème.

Concernant les mesures chronométriques, nous avons choisi d'examiner la même proportion de pauses pour chaque texte (les 20 % de pauses les plus longues), ce qui nous conduit à nous référer à un seuil différent pour chaque texte. Les résultats montrent que pour les 3 groupes le seuil moyen pour les

pauses longues est à plus de 1 seconde, qui est le seuil le plus communément utilisé pour les adultes dans la littérature (Schilperoord, 2002). Il est à noter également que le seuil moyen diminue à mesure que le niveau scolaire augmente. Dans le cadre de la théorie capacitaire (McCutchen, 1996), cette diminution pourrait être mise en relation avec des changements au niveau des couts cognitifs : par exemple, au fil de l'expérience certaines opérations gagnent en automaticité.

L'usage des marques graphiques de segmentation, nous l'avons vu, présente des écarts par rapport à la norme dans nos textes. Nous avons codé l'absence de virgule et de connecteur à une frontière de PS, l'absence de majuscule en début de PG ou l'absence de point en fin de PG. Les résultats indiquent une baisse moyenne des écarts par rapport à la norme concernant les deux premiers indicateurs (virgules, majuscules) avec le niveau scolaire. L'absence de point est l'écart par rapport à la norme le moins présent dans nos textes et ne fait de différence qu'entre les CM2 et les élèves plus âgés (5ème et 3ème). Cela rejoint des études qui ont montré que la ponctuation des phrases était mal maitrisée en 6ème, même si les élèves savent, d'un point de vue théorique, qu'une phrase graphique est limitée par une majuscule et un point (Fauchart, 2001). Toutefois, l'examen des moyennes masque l'existence de profils de scripteurs selon ces dimensions. Nous avons ainsi repéré des scripteurs qui avaient globalement peu de difficultés visibles dans les textes (Profil A3) dont seulement 10 sur 72 étaient en CM2 ; une trentaine de scripteurs ont été identifiés comme ayant des difficultés quant aux virgules et connecteurs en frontière de PS et enfin une vingtaine d'élèves marque peu les majuscules en début de PG. Dans ces deux derniers profils se trouvent essentiellement des élèves de CM2.

En ce qui concerne la longueur des pauses en fonction du type de frontière syntaxique, les résultats suggèrent que le travail de planification est mieux reporté sur des frontières fortes en 5ème et en 3ème qu'en CM2. Une pause longue suggère en effet un cout cognitif important, pouvant correspondre par exemple à une activité de planification (Foulin, 1998; Passerault, 1991) ou de relecture de ce qui vient d'être écrit. La proportion de frontières de PG correspondant à des pauses longues étant plus importante en 3ème qu'en 5ème, et plus en 5ème qu'en CM2, cela suggère une plus forte adéquation de la planification à des frontières fortes. Les frontières de PS correspondant à des pauses longues connaissent sensiblement la même répartition à travers les niveaux scolaires, ce qui serait conforme à nos attentes, les frontières de PS étant des frontières fortes. Enfin les frontières de C marquées par des pauses longues sont en proportion moindre en 3ème et cela conforterait notre interprétation, les frontières de C étant moins fortes. Les profils de scripteurs qui émergent des données sont d'une part des scripteurs qui planifieraient en frontière de C ou de PS, stratégie fréquente chez les élèves de CM2 ; des scripteurs qui planifieraient en frontière de PS ou PG, stratégie fréquente chez les élèves de 5ème et de 3ème ; ou enfin, des scripteurs qui planifieraient essentiellement en frontière de PG. Outre une normalisation de la segmentation (Paolacci & Favart, 2010) les élèves tendent donc vers un profil chronométrique plus proche de celui des adultes, avec des planifications plutôt au niveau des PS voire PG, ce qui corrobore l'étude menée précédemment sur les unités de planification. Certes, les longues pauses peuvent parfois servir à relire ce qui a été écrit

plutôt qu'à planifier la suite, mais le choix de s'arrêter en frontière d'unité syntaxique implique que cette unité peut être considérée comme une unité de production.

En mettant en lien les profils chronométriques et les profils quant aux écarts par rapport à la norme, on remarque que les élèves qui répondent le mieux à la norme entrent dans des profils qui ont des proportions importantes de pauses longues aux frontières de PG et de PS. Plus particulièrement, un nombre assez important de personnes dont les textes ont une proportion élevée de pauses longues en frontières de C et de PS (profil B1) font beaucoup d'erreurs soit de virgule, soit de majuscule (profils A1 et A2) ; sur ces 15 élèves, 14 sont en CM2. À l'inverse, parmi ceux qui ont peu de problèmes de ponctuation (profil A3), une très grande partie (66 élèves) font des pauses longues en début de PS et de PG (profil B2) ; parmi eux 19 sont en 5ème et 24 sont en 3ème. L'interprétation du profil de scripteur faisant des pauses longues en frontières de PG (profil B3) est plus difficile : le nombre d'élèves de ce profil est assez stable en fonction des niveaux. Cependant, alors que 10 élèves sur 11 en 3ème ont peu de problèmes de ponctuation, les élèves de CM2 ayant ce profil ont majoritairement des problèmes avec les virgules ou les majuscules. Une limite importante pour cette interprétation est le fait que des textes de longueurs différentes composent notre corpus. La prise en compte des phénomènes de façon proportionnelle ne nous garantit en effet pas que leur gestion dans le temps puisse être comparable. Par exemple, Fayol (1989) montre que, dans les narrations, la ponctuation se manifeste aux frontières des épisodes, délimitant ainsi des parties dans le texte, dont certaines sont « sur-ponctuées » (le résumé initial, et la fin du récit) et d'autres ne présentent pas de ponctuation (par exemple pas de ponctuation de la complication jusqu'au dénouement). La proportion de texte constituant les épisodes pourrait ainsi affecter nos résultats.

2.5. Conclusion

Les analyses menées dans cette étude confirment la pertinence de recourir à la segmentation en « phrases syntaxiques » (Paolacci & Rossi Gensane, 2014) afin de mieux rendre compte des comportements des scripteurs. En effet, la PS semble être l'unité de planification de la majorité des scripteurs à partir de la classe de 5ème. Cette unité A en outre l'avantage d'être définie uniquement en termes syntaxiques, contrairement à la « phrase », dont la définition mêle des critères sémantiques, prosodiques, énonciatifs, et dont la segmentation prend souvent davantage de sens lorsqu'elle est considérée au sein du texte, et non « en soi » (Rondelli, 2013). Une fois les phrases syntaxiques délimitées, il serait possible de discuter avec les élèves de la meilleure marque de segmentation à mettre : en effet, un enchainement de PS apparait toujours comme non canonique s'il n'y a ni marque de ponctuation, ni connecteur logique ; or, les problèmes de marquage des frontières PS restent présents même pour certains élèves de 3ème. En discutant de la meilleure marque à mettre à cette frontière, il ne s'agirait plus d'aborder la ponctuation comme marque de segmentation, mais plutôt comme marque de liaison. En effet, cela pourrait ouvrir la discussion sur la diversité des marques de ponctuation en tant

que reflet du degré de liaison entre les unités, par exemple suivant l'organisation hiérarchique du texte (Fayol, Carré & Simon-Thibult, 2014), ce qui permettrait aussi de développer l'usage du point-virgule, souvent rare dans les productions d'élèves. Par ailleurs, les unités sont également liées suivant le sémantisme : dans ce cas, l'usage d'un connecteur approprié peut être davantage informatif qu'une virgule ou un point ; là encore, des marques de ponctuation plus rarement utilisées, comme les deux points, pourraient être abordées à travers cette fonction. Dès lors, l'absence de phrase graphique dans un texte devrait poser question à un élève : cela signifierait que son texte manque d'articulation, si tout est uniquement relié par des virgules ou la conjonction « et ». Un travail d'oralisation pourrait alors permettre d'identifier où, spontanément, l'enfant marquerait une pause un peu plus longue, ce qui correspondrait à une frontière de phrase graphique. Peu à peu, la phrase graphique pourrait ainsi prendre sens, se rapprochant de la notion de « période » (en tant qu'unité prosodique), avec en son sein une ou plusieurs phrases syntaxiques correctement reliées les unes aux autres. Nous pouvons supposer que ce n'est que plus tard que, la phrase graphique ayant pris sens pour le scripteur, ce dernier sera en mesure de planifier davantage au niveau de la phrase graphique que de la phrase syntaxique, adoptant un comportement plus proche de celui des adultes.

Cette étude gagnerait à intégrer la question des connecteurs ainsi que celle de la disposition graphique des clauses (e.g. regroupements en paragraphes). Enfin, nous disposons pour chaque élève d'un texte narratif et d'un texte expositif qui ont été considérés ensemble dans cette étude et qui pourraient, considérés séparément, apporter un autre éclairage sur la segmentation.

3. Discussion générale

3.1. Vers une planification au-delà de la clause

Ces deux études abordent de manière complémentaire la question des unités de production pour les scripteurs de 10 à 15 ans. Dans la première étude, les données chronométriques sont utilisées comme indices des moments de planification. Il ressort que les plus jeunes planifient plutôt clause par clause, de manière similaire aux adultes traduisant un texte (Immonen, 2006; Immonen & Mäkisalo, 2010). Par la suite les scripteurs planifient plusieurs structures à un niveau supra-clausal : par exemple, les élèves de 5ème et de 3ème semblent planifier les clauses enchâssées pendant l'écriture de la clause précédente ; dans le cas des subordonnées en début d'unité T, les élèves de 3ème font une pause initiale très longue, sans doute pour planifier à la fois la subordonnée et la matrice. En outre, les élèves de 3ème s'arrêtent très longtemps avant les unités T, alors que les processus de plus bas-niveau sont davantage automatisés : cela suggère que durant la pause entre deux unités T, voire deux phrases puisque les phrases sont précédées par des pauses plus longues que les unités T, le scripteur de 3ème est engagé dans une activité plus couteuse cognitivement. Nous avons à de nombreuses reprises évoqué l'activité de planification, mais il peut également s'agir de relecture du texte ; cependant, même s'il s'agit de relecture, le fait de marquer un arrêt peut indiquer que l'unité écrite est considérée comme close. Cette augmentation du cout peut être due à un processus durant lequel le scripteur s'engage plus profondément que lorsqu'il est plus jeune : par exemple, une recherche dans la MLT plus longue car prenant en compte davantage de connaissances, ou un processus de révision plus attentif aux différents niveaux du texte qui sont à vérifier. Selon nos hypothèses, ce cout plus élevé peut également être dû à un processus porté sur une plus grande unité : si la planification (ou la relecture) ne porte pas que sur une clause mais sur un ensemble de clauses entretenant un lien de dépendance syntaxique, la pause sera sans doute plus longue.

Cependant, avant de pouvoir affirmer que l'élève est en mesure de planifier ensemble plusieurs clauses, il faut s'assurer que ces unités supra-clausales ont un sens pour lui. Si l'on part du principe que l'enfant plus jeune planifie clause par clause, il faut observer comment il arrive à lier ces clauses entre elles pour en faire des unités T complexes, et ponctue correctement son texte pour présenter des phrases graphiques. La deuxième étude fait apparaitre que les élèves de CM2 sont ceux qui ont la plus forte proportion de pauses longues entre les clauses, et qu'ils présentent le plus de problèmes de ponctuation. Surtout parmi les jeunes scripteurs, les cas où plusieurs clauses entretiennent un lien de dépendance syntaxiques sont gérés différemment que pour les scripteurs plus expérimentés : ils s'arrêtent longtemps avant les clauses à l'intérieur des unités T, et ont du mal à lier correctement deux unités T, celles-ci n'étant parfois reliées ni par un connecteur, ni par une virgule. À l'inverse, les élèves ayant peu de problèmes de ponctuation font généralement peu de pauses longues entre les clauses lorsque celles-ci sont à l'intérieur d'une unité T. Il est ainsi possible qu'une bonne compréhension de la manière de

grouper les clauses en unités plus larges en indiquant correctement les liens entre ces différentes unités facilite une planification plus globale. Or, des difficultés pour lier correctement les unités T étant toujours présentes en 3ème, il apparait important de poursuivre les réflexions sur la manière d'enseigner le lien entre les clauses, souvent abordé à travers la notion de « segmentation en phrase ».

3.2. Perspectives de recherche et limites du corpus

Ces deux études suggèrent donc qu'avec l'âge, les scripteurs parviennent à planifier ensemble les clauses qui ont un lien fort de dépendance syntaxique. La question que l'on souhaiterait se poser alors est : quel est le contenu exact de ce qui est planifié aux différents moments de l'écriture (pendant la pause en début d'unité T, pendant la clause matrice, etc.) ? Le problème est que les études qui viennent d'être présentées soulèvent de nombreuses limites quant aux recherches que l'on pourrait mener sur ce corpus.

Tout d'abord, l'étude des profils montre que les groupes n'étaient pas complètement homogènes concernant la maitrise de la ponctuation et la longueur des pauses en fonction des emplacements syntaxiques, même si certains profils étaient privilégiés par les élèves d'un certain niveau scolaire : il faudrait alors s'interroger sur la pertinence d'étudier plutôt le développement des processus rédactionnels en fonction de la maitrise de certaines compétences linguistiques ou cognitives, et non simplement en fonction des niveaux scolaires, afin de mieux caractériser la progression des élèves. Cela nécessiterait d'avoir à notre disposition davantage de données concernant les élèves (comme l'empan de leur MDT, leur fluence d'écriture), que nous n'avons pas en notre possession pour ce corpus.

Par ailleurs, le corpus Reflex est relativement hétérogène concernant le contenu, puisque deux types de textes très différents, un expositif et un narratif, sont produits. Le type de contenu linguistique de ces textes est différent et évolue avec l'âge (Berman, 2004a), tout comme le type de stratégie de planification utilisé (Bereiter & Scardamalia, 1987). Dès lors, il faudrait étudier séparément les deux types de texte, ce qui permettrait de faire ressortir ce qui est dû à une stratégie contextuelle de ce qui relève d'une compétence générale.

Le dernier problème quant à l'utilisation de ce corpus pour étudier le contenu de la planification aux différents moments de l'écriture d'une phrase provient de la faible quantité de données utilisables. En effet, une telle analyse nécessite de travailler sur des portions de texte qui ne comportent pas de séquences de révision, or la plupart des phrases de ce corpus contiennent des révisions : par conséquent, même si des résultats étaient obtenus, ils seraient sans doute peu fiables pour rendre compte du processus d'écriture de manière générale.

Comme la question principale de notre travail concerne le contenu de la planification au niveau de la phrase dans un texte écrit, les nombreuses limites soulevées nécessitent l'utilisation d'un autre corpus. Cette question précise n'ayant jamais été étudiée chez les adultes, il semble que l'utilisation d'un corpus

plus homogène produit par une population de scripteurs adultes soit davantage pertinente pour répondre avec fiabilité à cette problématique. Cela nous a conduits à poursuivre nos analyses sur le corpus « Dimitri », dont les détails de la collecte de données ont été donnés dans le chapitre « Méthodologie générale ». En approfondissant notre analyse à partir d'un corpus adulte, nous espérons mieux saisir les rapports entre temporalité de l'écriture et contenu linguistique afin de faire émerger les unités de production d'un texte écrit. De nouvelles perspectives pourront alors être évoquées pour poursuivre une réflexion sur le développement de ces processus rédactionnels.

Chapitre 6 – Données chronométriques du corpus adulte

Le corpus « Dimitri » a été produit par trente scripteurs adultes. Ces derniers avaient pour consigne de rédiger la suite d'une histoire intitulée *L'évasion* ; ce texte devait faire entre une demi-page et une page. Le protocole expérimental devrait permettre d'obtenir un corpus relativement homogène au niveau du contenu linguistique et du type de participants ayant pris part à cette étude. Ainsi, il sera possible d'explorer en détail la gestion temporelle des processus rédactionnels, notamment afin de définir à quel moment les processus de micro-planification et de formulation prennent place.

Deux analyses préliminaires, qui font l'objet de ce chapitre, sont nécessaires afin d'explorer le lien entre temporalité de l'écriture et caractéristiques linguistiques du texte. D'une part, une description générale des données chronométriques permettra de mieux comprendre les liens entre les différentes variables chronométriques et leur évolution au cours du texte, et exposera également les différences pouvant être notées entre les scripteurs. D'autre part, nous nous intéresserons par la suite plus particulièrement à la gestion temporelle des processus au niveau local, en mettant en lien les données chronométriques avec le contenu des différentes unités linguistiques. Il convient donc de justifier tout d'abord le recours à certaines unités de segmentation définies par des critères linguistiques, ce qui sera l'objet de la deuxième partie de ce chapitre.

1. Description générale des données chronométriques

La description chronométrique a pour but de caractériser le comportement des scripteurs quant à la durée de leurs pauses en début de mot, pauses intra-mots et temps d'exécution. Ce comportement sera observé de manière générale, mais également en fonction des parties du texte et de la localisation syntaxique des séquences. Une indication particulière concernant la gestion temporelle de l'écriture sera fournie : la manière de s'arrêter avant et après les signes de ponctuation.

Le comportement chronométrique est tout d'abord indiqué de manière générale sur le corpus, à partir de la moyenne des trente scripteurs ; l'écart-type est donné entre parenthèses. Il sera également indiqué s'il existe des corrélations entre les variables observées. Afin d'établir des « profils » de scripteurs pour chaque groupe de variables examiné, nous avons utilisé le test « classification hiérarchique » sous SPSS. Nous avons ensuite vérifié l'établissement de ces profils grâce à l'analyse discriminante, qui permet en outre d'indiquer quelles variables diffèrent d'un profil à l'autre. Un graphique récapitule les caractéristiques de chaque profil.

Cette première partie est plutôt descriptive, mais nous pouvons cependant formuler quelques hypothèses générales, notamment grâce aux études présentées dans le cadre théorique.

- Hyp. 1 : Les processus cognitifs engagés diffèrent suivant l'emplacement syntaxique (« I », « C », « A », « P »[54]). Ces résultats viendraient corroborer des résultats déjà observés sur des productions écrites et orales ; ils seraient affinés ici par la distinction entre unité A et phrase.
    - Hyp 1a : Le temps passé avant et après le signe de ponctuation diffère en fonction de l'emplacement syntaxique : plus l'emplacement syntaxique est élevé dans la hiérarchie, plus la pause se situera après le signe de ponctuation (cf Foulin et al., 1989).
    - Hyp. 1b : La durée de la pause initiale diffère en fonction de l'emplacement syntaxique : plus l'emplacement syntaxique est élevé dans la hiérarchie, plus la pause initiale sera longue (e.g. Chanquoy et al., 1996; Foulin, 1998; Immonen, 2006; Sanders & Schilperoord, 2006).
- Hyp. 2 : Tous les scripteurs sont soumis à de nombreuses contraintes (cognitives, linguistiques, etc.) lorsqu'ils rédigent un texte. La surcharge cognitive va contraindre le scripteur à ralentir son rythme d'écriture (Brown et al., 1988), mais cela peut se faire de différentes manières : par des pauses initiales plus longues, davantage de pauses intra-mots ou des temps d'exécutions plus longs. Si la

[54] « P » désigne une pause inter-phrases ; « A » une pause en début d'unité A qui ne soit pas une nouvelle phrase ; « C » une pause en début de clause qui ne soit pat une nouvelle unité A ; « I » une pause avant un mot situé à l'intérieur d'une clause. Dans cette étude, l'unité A est préférée à l'unité T car elle sera plus opératoire lors de la description syntaxique et sémantique des phrases.

surcharge cognitive est la même pour tous les étudiants, un équilibre devrait se créer entre durée des pauses initiales, des pauses intra-mots et du temps d'exécution.

- Hyp. 2a : Plus la pause initiale est longue, plus les pauses intra-mots et le temps d'exécution sont courts.
- Hyp. 2b : Il y a une corrélation négative entre les pauses intra-mots et le temps d'exécution.

- Hyp.3 : Des profils de scripteurs concernant la gestion temporelle globale de l'écriture peuvent être identifiés. Ainsi, certains vont faire davantage de longues pauses de macro-planification alors que d'autres vont planifier et relire leur texte au fur et à mesure (Doquet, 2006) ; en outre, la quantité de révisions peut être liée à la longueur des pauses initiales (Baaijen et al., 2012).
  - Hyp. 3a : Des scripteurs ont un rythme d'écriture plutôt irrégulier au long du texte, avec notamment des pauses très longues, alors que d'autres font des pauses plus courtes et sont plus stables tout au long du texte.
  - Hyp. 3b : Les scripteurs faisant moins de pauses très longues font davantage de révisions au cours de leur texte.

1.1. Ponctuation

1.1.1. Description générale

Comme indiqué lors du codage, lorsqu'une unité était précédée par un signe de ponctuation nous avons fait le choix de considérer comme « pause initiale » la somme de la pause initiale du signe de ponctuation et de la pause initiale du premier mot de l'unité[55]. La distribution du temps de pause avant et après le signe de ponctuation peut dépendre des personnes, ainsi que de la localisation de la ponctuation. Le Tableau 16 indique le nombre de ponctuations dans le texte et la répartition du temps de pause à ces endroits, cette dernière étant illustrée par la Figure 24.

En moyenne, il y avait 31 signes de ponctuation par texte. Comme le montre la Figure 23, plus de la moitié de la ponctuation marquait des frontières de phrases (« avant « P » »). Les autres marques étaient des virgules, placées soit en début de clause à l'intérieur d'une phrase (« avant « A » ou « C » »), soit à l'intérieur des clauses (« avant « I » »).

De manière générale, sur le temps total de pause entre le mot avant et après le signe de ponctuation, seul 27% du temps de pause était pris avant la ponctuation. Cependant on observe une nette différence suivant l'emplacement de la ponctuation : à l'intérieur des phrases, les pauses sont généralement plus longues avant la ponctuation (67% avant « A » ou « C », 57% avant « I ») qu'après le signe de

[55] Nous n'avons pas pris en compte le temps d'écriture du signe lui-même, qui est toujours très court : on peut estimer qu'il s'agit d'un temps nécessaire d'un point de vue graphique, et trop court pour que des processus cognitifs soient engagés.

ponctuation ; à l'inverse, 84% du temps de pause se situe après le signe de ponctuation séparant deux phrases.

	Total	avant « P »	avant « A » ou « C »	avant « I »
nombre de ponctuations par texte	31,00 (8,51)	17,83 (5,86)	7,07 (4,18)	6,37 (3,54)
% du temps de pause avant la ponctuation	27% (12%)	16% (10%)	67% (24%)	57% (23%)

Tableau 16. Nombre de signes de ponctuation et répartition du temps de pause en fonction de l'emplacement syntaxique (moyennes et écart-types)

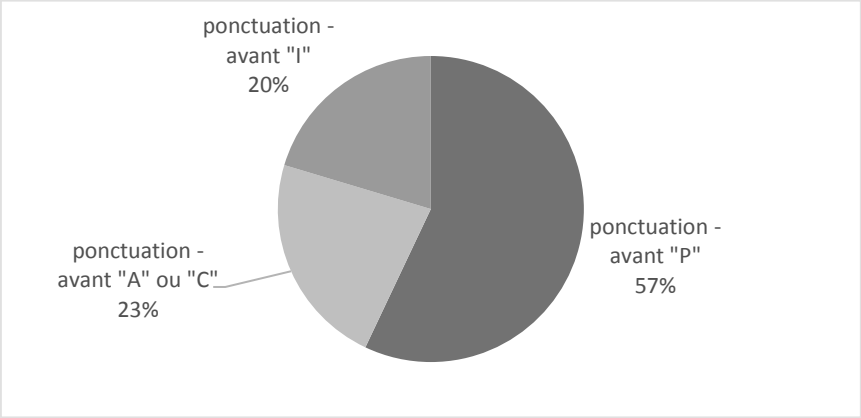


Figure 23. Répartition de la ponctuation en fonction de l'emplacement syntaxique (en %)

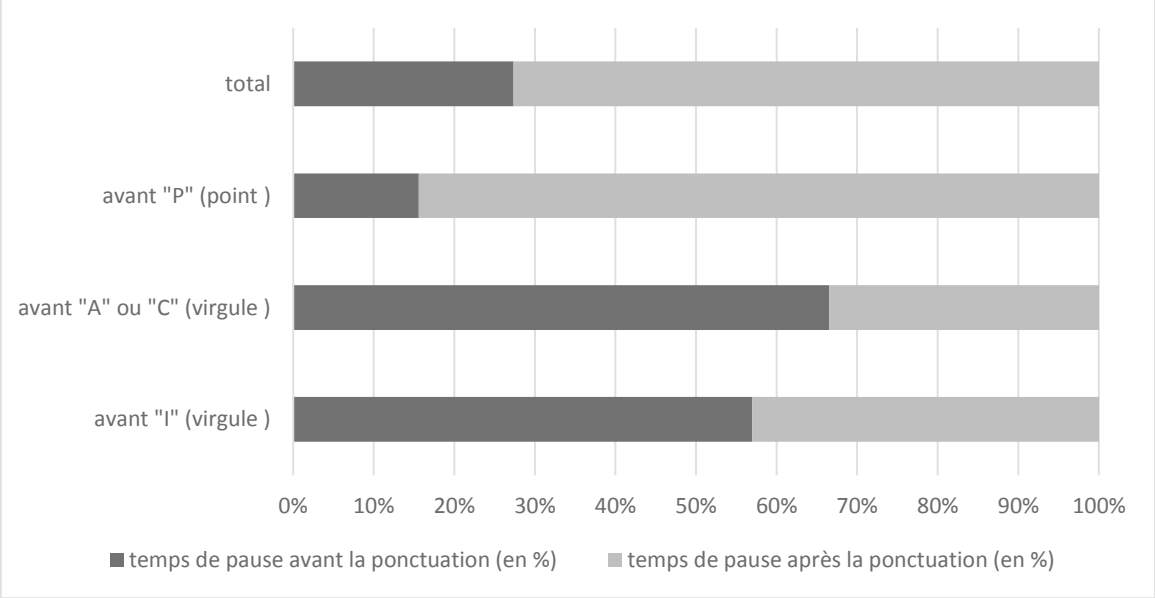


Figure 24. Distribution du temps de pause avant et après les signes de ponctuation en fonction de l'emplacement syntaxique (moyennes en %)

Entre ces variables, il n'y a aucune corrélation entre la quantité de signes de ponctuation et la part de pause passée avant ou après l'écriture du signe. Il y a en revanche une corrélation entre le nombre total de signes de ponctuation et le nombre de ponctuations avant « P » (r=0,72), avant « A » ou « C » (r=0,51) et avant « I » (r=0,62), ce qui est logique puisque les ponctuations entre phrases représentent

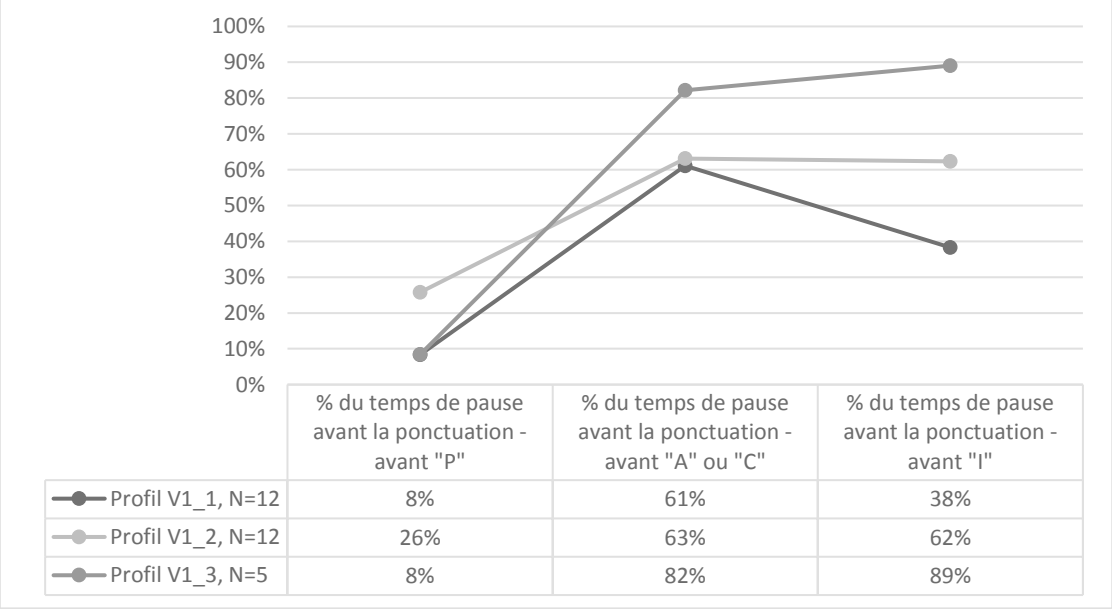
la plus grande part des signes de ponctuation. Il n'est alors pas étonnant que le pourcentage du temps de pause avant la moyenne globale (« total ») soit fortement corrélé à celui avant « P » (r=0,80) et significativement corrélé à celui avant « A » ou « C ».

1.1.2. Profils

Nous avons utilisé le test « classification hiérarchique » afin d'établir des profils. Nous avons entré comme variables celles concernant la proportion de temps avant le signe de ponctuation, dans le cas des trois emplacements syntaxiques (« P », « A / C », « I »). Les variables de la moyenne globale (« total ») étant fortement corrélées à celle avant « P » (r>0,7), nous ne les avons pas insérées.

Trois profils ont été retenus ; ils sont correctement classés par l'analyse discriminante. Les tests ANOVAs révèlent que la différence entre les groupes est significative pour le pourcentage de temps avant « P » et avant « I », mais pas avant « A » ou « C ». Ainsi, le profil 2 se caractérise par un temps de pause avant la ponctuation entre phrases plus élevé que les autres profils (25%), et un temps moyen avant les ponctuations à l'intérieur des clauses (62%). Les participants des profils 1 et 3 s'arrêtent peu avant la ponctuation en fin de phrase, et ont un comportement différent pour la ponctuation avant « I » : les participants du profil 1 s'arrêtent assez peu avant (38%), alors que dans le profil 3 presque tout le temps de pause (89%) se situe avant le signe de ponctuation. Ces profils sont résumés dans le Tableau 17 et illustrés par la Figure 25.

L'hypothèse 1a est donc partiellement validée, puisque les moyennes indiquent que le pourcentage de temps de pause avant la ponctuation est plus faible avant « P » qu'avant « A / C » et « I », mais le comportement avant « I » diffère fortement suivant les profils.



	% du temps de pause avant la ponctuation - avant "P"	% du temps de pause avant la ponctuation - avant "A" ou "C"	% du temps de pause avant la ponctuation - avant "I"
Profil V1_1, N=12	8%	61%	38%
Profil V1_2, N=12	26%	63%	62%
Profil V1_3, N=5	8%	82%	89%

Figure 25. Profils liés à la ponctuation (moyennes en %)

V1[56]	Ponctuation	Profil_V1_1	N=12	Part de pause faible avant les ponctuations précédant « P » et « I »
		Profil_V1_2	N=12	Part de pause moyenne avant les ponctuations précédant « P » et « I »
		Profil_V1_3	N=5	Part de pause faible avant les ponctuations précédant « P », part de pause très important avant les ponctuations précédant « I »

Tableau 17. Ponctuation - Caractéristiques des profils (V1)

1.2. Gestion des pauses et des temps d'exécution

1.2.1. Moyenne des pauses et des temps d'exécution

1.2.1.1. Description générale

Diverses variables ont été relevées et moyennées sur l'ensemble du texte, pour chaque sujet : la pause de pré-écriture, les pauses initiales des mots en fonction de leur emplacement syntaxique, les pauses intra-mots et le temps d'exécution. Le Tableau 18 indique la moyenne de ces variables par texte ; plus spécifiquement, la durée de la pause initiale en fonction de l'emplacement syntaxique est illustrée dans la Figure 26.

		Moyenne par texte
Temps total d'écriture		1048529,37 (338307,16) = 17 min et 28 secondes
Pause pré-écriture		74454,73 (51690,87) = 1 min et 14 sec
Pauses initiales (en ms) - total[57]		1534,19 (1643,20)
	pause avant « P »	9444,84 (5634,45)
	pause avant « C »	3565,82 (2672,13)
	pause avant « A »	2985,83 (2851,21)
	pause avant « I »	951,85 (450,07)
Pauses intra-mot (en ms/car)		141,33 (60,04)
Exécution (en ms/car)		280,35 (55,78)

Tableau 18. Durées des pauses et des temps d'exécution (moyennes en ms et écart-types)

[56] Un tableau récapitulatif indiquera les profils déterminés pour chaque groupe de variables. Ici, il s'agit du groupe « V1 », concernant la ponctuation, pour lequel 3 profils émergent.

[57] La variable « pause initiale – total » réfère aux pauses initiales tout emplacement syntaxique confondu ; pour rappel, les « pauses initiales » cumulent toujours la pause avant et après le signe de ponctuation précédant le mot, le cas échéant.

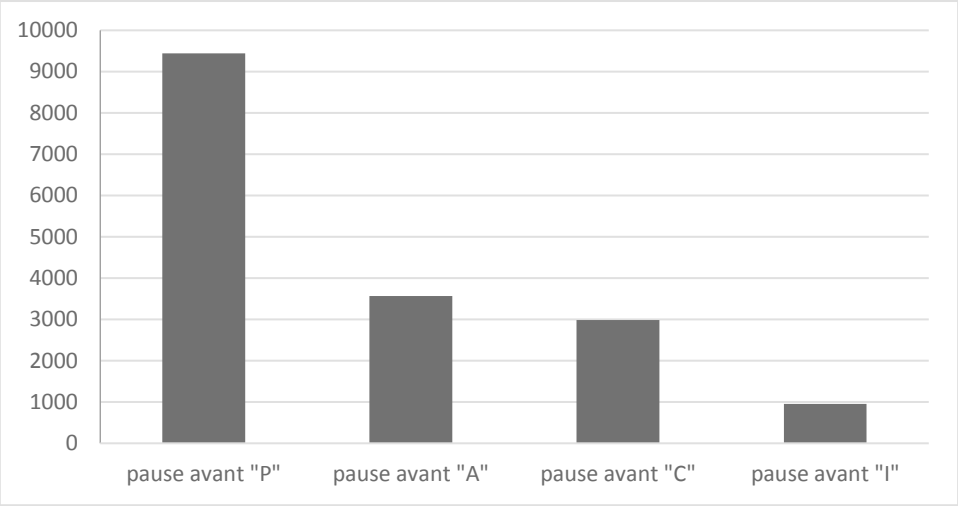


Figure 26. Durée de la pause initiale en fonction de l'emplacement syntaxique (moyennes en ms)

La moyenne globale de la durée des pauses initiales (« total ») est fortement reliée ($0,72<r<0,92$) à la durée des pauses initiales de chaque emplacement syntaxique. En outre, il y a des corrélations significatives positives ($0,55<r<0,68$) entre toutes les pauses initiales avant chaque emplacement syntaxique : plus la pause initiale est longue avant un emplacement, plus elle est longue devant les autres emplacements ; seule la corrélation entre la durée des pauses avant « A » et celles avant « C » n'est pas significative. La pause de pré-écriture est significativement corrélée avec les pauses initiales des emplacements syntaxiques « C » ($r=0,50$) et « I » ($r=0,43$). Le temps total d'écriture, que la pause de pré-écriture soit ou non inclue, est quant à lui significativement corrélé à toutes les durées de pause initiale ($0,39<r<0,62$), hormis la pause avant « A ». Enfin, la seule corrélation significative concernant les pauses intra-mots ou le temps d'exécution consiste en une corrélation positive entre les pauses intra-mots et les pauses avant « I » ($r=0,39$). Par ailleurs, plus le temps d'écriture est long, plus la pause de pré-écriture est longue ($r=0,62$), la pause de pré-écriture représentant en moyenne 7% du temps total d'écriture.

1.2.1.2. Profils

La pause de « pré-écriture » correspond normalement au temps entre la fin de la lecture de la consigne et l'écriture du premier mot : le scripteur devait tracer une croix lorsqu'il finissait de lire l'amorce (titre du texte et début de l'histoire) et commençait à réfléchir, ce qui correspondait au début de ce temps de « pré-écriture ». Cette donnée chronométrique se distingue des autres, qui correspondent à la tâche d'écriture tout au long du texte. C'est pourquoi, afin de ne pas fausser les analyses, nous la gardons de manière indépendante, et les profils ne seront établis qu'à partir des autres variables (pauses initiales, pauses intra-mots et temps d'exécution). Nous n'avons pas entré les moyennes des pauses avant chaque emplacement syntaxique car celles-ci étaient fortement corrélées à la moyenne globale des pauses initiales (« total ») ($r>0,72$), et parce que la section suivante détaillera davantage la durée des pauses pour chacun de ces emplacements.

La répartition finale choisie est une classification en 3 profils. L'analyse discriminante classe correctement les individus dans ces 3 profils à partir des 3 variables entrées. Des différences significatives existent pour ces 3 variables.

Les caractéristiques de ces profils sont résumées dans le Tableau 20 et illustrées dans la Figure 27. Cette figure est obtenue à partir des valeurs centrées-réduites, à cause de la différence d'échelle entre les variables, c'est pourquoi nous avons ajouté le Tableau 19 pour indiquer les moyennes de chaque profil. Le graphique montre certaines interactions entre la durée des pauses intra-mots et le temps d'exécution : le profil 1, qui a le temps d'exécution le plus long, est celui dont les pauses intra-mots sont les plus courtes, alors que les scripteurs du profil 2 ont un temps d'exécution court mais font des pauses intra-mots très longues. Ces profils se distinguent entre eux par la durée des pauses initiales : celles-ci sont en moyenne longues pour les scripteurs du profil 3, et courtes pour ceux du profil 1.

	profil 1, N= 11	profil 2, N= 12	profil 3, N= 7
pauses initiales (ms)	947,93	1684,67	2664,66
pauses intra-mot (ms/car)	92,49	195,64	124,99
exécution (ms/car)	306,20	251,53	289,16

Tableau 19. Moyenne des pauses et du temps d'exécution par profil (en ms)

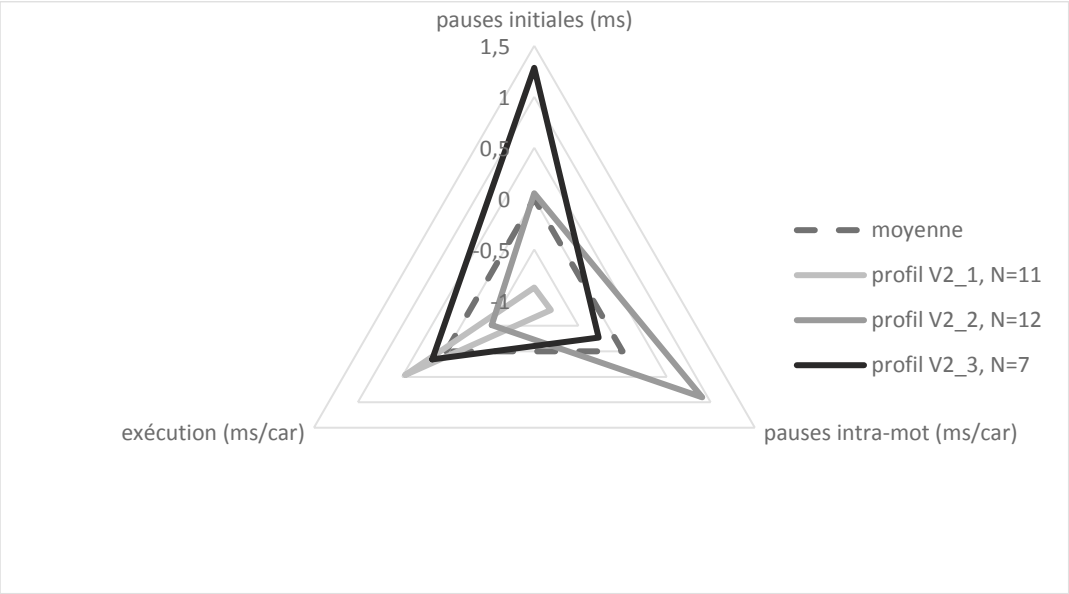


Figure 27. Profils liés à la moyenne des pauses et du temps d'exécution

V2	Durée des pauses et de l'exécution	Profil_V2_1	N=11	Pauses initiales courtes, pauses intra-mots courtes et temps d'exécution élevé
		Profil_V2_2	N=12	Pauses initiales moyennes, pauses intra-mots longues et temps d'exécution faible
		Profi_V2_3	N=7	Pauses initiales longues, pauses intra-mots moyennes et temps d'exécution élevé

Tableau 20. Moyenne des temps de pause et d'exécution - Caractéristiques des profils (V2)

1.2.2. Catégorie des pauses

1.2.2.1. Description générale

Les pauses ont été catégorisées suivant leur durée. Les classes de durée étaient : 0-100 ms ; 100-250 ms ; 250-500 ms ; 500-1000 ms ; 1000-2000 ms ; 2000-10000 ms ; plus de 10000 ms. Pour chaque emplacement (avant « P », avant « A », avant « C » et avant « I »), il a été calculé le pourcentage de pauses dans chaque catégorie, ces informations étant données dans le Tableau 21.

	0-100	100-250	250-500	500-1000	1000-2000	2000-10000	10000-plus
avant « P »	0% (0%)	1% (3%)	3% (5%)	6% (6%)	13% (12%)	43% (14%)	34% (19%)
avant « A »	0% (2%)	3% (10%)	17% (17%)	22% (22%)	15% (12%)	33% (22%)	10% (14%)
avant « C »	1% (3%)	7% (10%)	27% (21%)	20% (12%)	15% (13%)	22% (15%)	8% (10%)
avant « I »	4% (4%)	31% (14%)	38% (14%)	13% (7%)	6% (2%)	6% (3%)	2% (2%)
total	4% (1%)	27% (12%)	35% (12%)	14% (7%)	7% (2%)	10% (4%)	4% (3%)

Tableau 21. Proportion de pauses par catégorie de durée (en ms) en fonction de l'emplacement syntaxique (moyennes en % et écart-types)

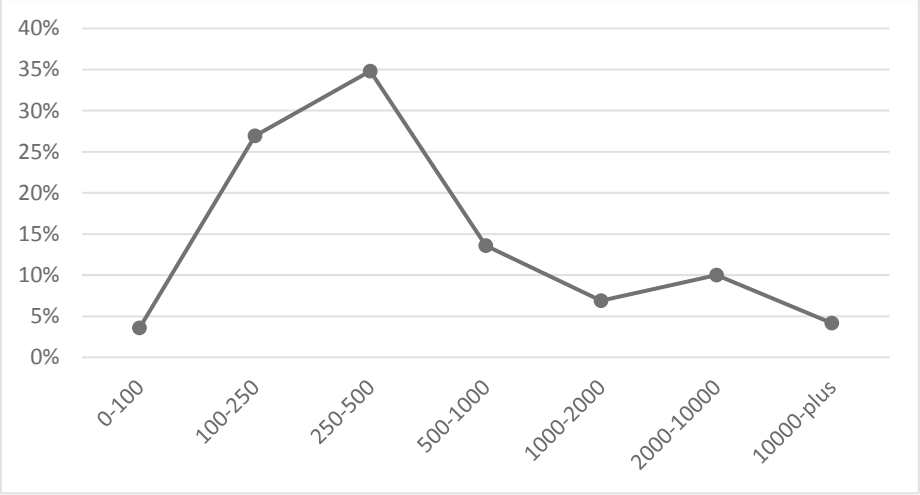


Figure 28. Proportion de pauses par catégorie de durées (durées en ms, moyennes en %)

De manière générale, 62% des pauses durent entre 100 et 500 ms. Nous pouvons cependant voir sur la Figure 28 un petit « pic » dans la catégorie 2000-10000 ; cette distribution pourrait s'expliquer en distinguant les pauses suivant l'emplacement syntaxique. Certaines catégories de durée ont donc été regroupées pour plus de clarté et elles ont été séparées en fonction de l'emplacement syntaxique, ce qui est illustré dans la Figure 29. On observe ainsi que près de 70% des pauses situées à l'intérieur des clauses, soit les pauses les plus nombreuses dans le texte, durent effectivement entre 100 et 500 ms. À l'inverse, 77% des pauses avant « P » durent plus de 2 secondes. Les pauses avant « C » et avant « A » durent principalement entre 100 ms et 10 secondes, avec davantage de pauses avant « C » entre 100 ms et 2 s, et davantage de pauses avant « A » entre 500 ms et 10 s.

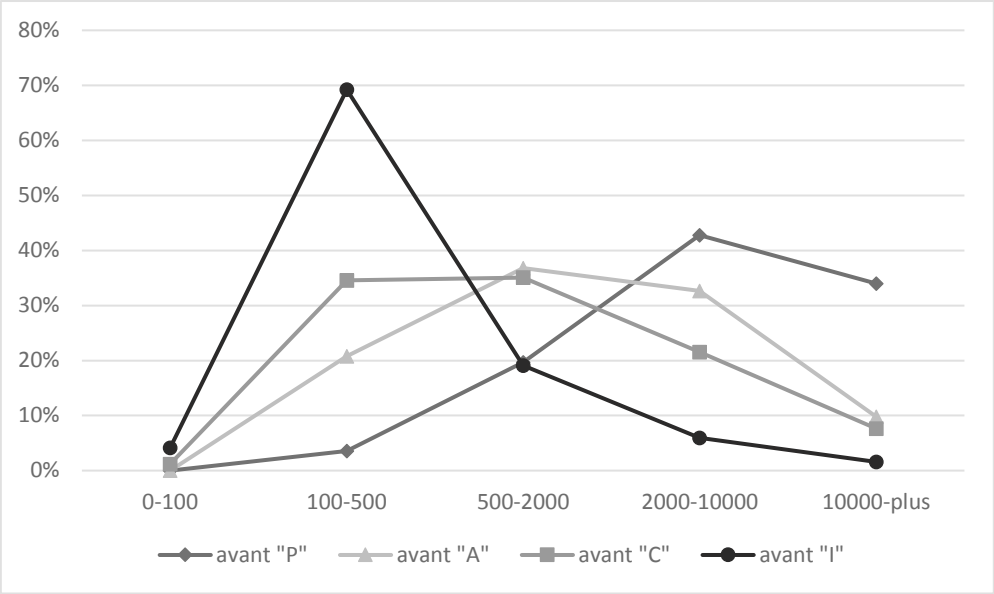


Figure 29. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique (durées en ms, moyennes en %)

Une matrice de corrélation fait ressortir un certain nombre de corrélations significatives entre toutes ces variables. Les corrélations fortes concernent notamment les catégories « totales » : celles-ci sont toujours très corrélées (0,81<r<0,99) avec les catégories équivalentes concernant les pauses avant « I », ce qui s'explique par la très forte proportion de pauses avant « I » dans le texte ; on peut également noter que la catégorie « % total – 10000-plus » est également très corrélée avec les catégories « % P 10000-plus » (r=0,81) et « % C 10000-plus » (r=0,78). Une autre forte corrélation peut être relevée : « % I 100-250 » est négativement corrélé à « % I 250-500 » (r=- 0,79).

1.2.2.2. Profils

Toutes les données (hormis le pourcentage « total » pour chaque catégorie) ont été entrées pour effectuer l'analyse. Deux profils émergent de cette classification. Les moyennes correspondant à chaque profil sont indiquées dans le Tableau 22.

	0-100	**100-250**	**250-500**	**500-1000**	**1000-2000**	**2000-10000**	**10000-plus**
Profil 1 (N=10)							
avant « P »	0%	0%	5%	12%	25%	42%	16%
avant « A »	0%	10%	27%	32%	16%	11%	5%
avant « C »	2%	15%	30%	20%	16%	14%	3%
avant « I »	7%	43%	31%	10%	5%	4%	1%
Profil 2 (N=20)							
avant « P »	0%	1%	1%	3%	8%	43%	43%
avant « A »	0%	0%	13%	16%	15%	44%	12%
avant « C »	1%	3%	26%	21%	14%	25%	10%
avant « I »	3%	25%	42%	15%	6%	7%	2%

Tableau 22. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique – Moyennes par profils (durée en ms, moyennes en %)

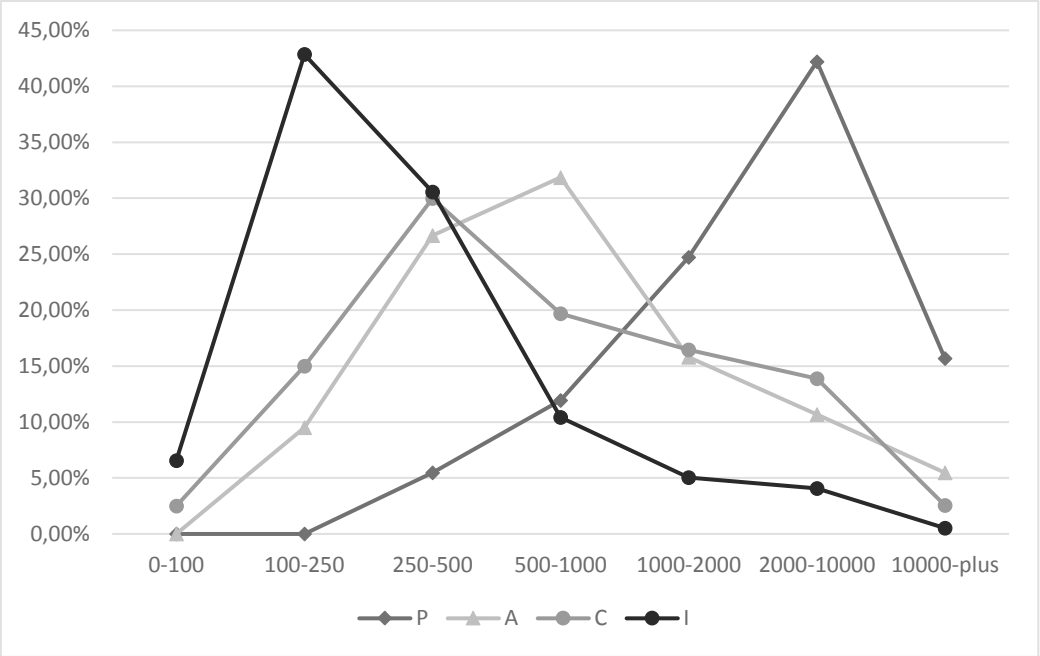


Figure 30. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique – Profil 1 (durées en ms, moyennes en %)

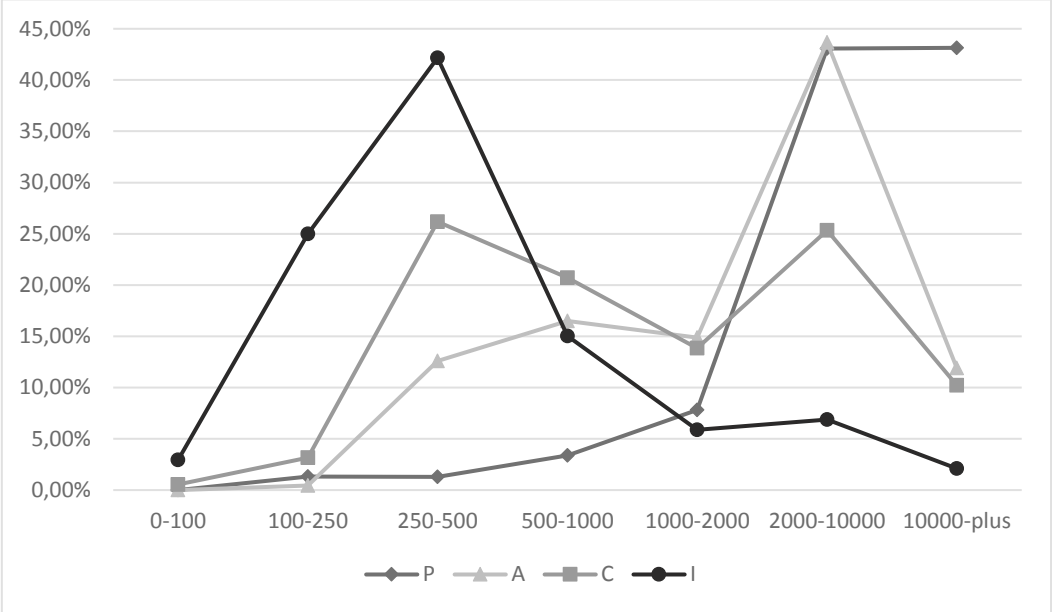


Figure 31. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique – Profil 2 (durées en ms, moyennes en %)

D'après l'analyse discriminante, qui est capable de reconstituer les groupes à 100%, 11 variables ne diffèrent pas significativement d'un groupe à l'autre : % P 100-250, % P 2000-10000, % A 500-1000, % A 1000-2000, % A 10000 +, % C 0-100, % C 250-500, % C 500-1000, % C 1000-2000, % I 500-1000, % I 1000-2000.

Le profil 1, illustré par la Figure 30, se distingue par la moitié des pauses avant « A » et avant « C » durant entre 250 et 1000 ms (avec un pic à 250-500 ms pour les pauses avant « C » et à 500-1000 ms avant « A ») ; les pauses avant « I » durent essentiellement entre 100 et 250 ms, les pauses avant « P »

entre 2 s et 10 s. Le profil 2, illustré par la Figure 31, a globalement des pauses longues : 40% des pauses « I » durent entre 250 et 500 ms, une part non négligeable (25%) des pauses C durent entre 2 et 10 s, les pauses « A » durent en grande partie entre 2 et 10 s, et près de 45 % des pauses « P » durent plus de 10 s.

V3	Pauses suivant emplacement	Profil_V3_1	N=10	pauses avant « I » plutôt entre 100 et 250 ms ; pauses avant « A » et « C » entre 250 et 1000 ms ; pauses avant « P » entre 2 et 10 s
		Profil_V3_2	N=20	Pauses plus longues

Tableau 23. Proportion de pauses par catégorie de durée en fonction de l'emplacement syntaxique - Caractéristiques des profils (V3) (durée en ms, moyennes en %)

Malgré les différences relevées ci-dessus, un schéma général reste constant, comme le montre la Figure 32 : les pauses « P » sont majoritairement supérieures à 2 secondes ; celles avant « I » généralement inférieures à 500 ms. Le comportement des pauses « A » et « C » est plus variable suivant les profils : le profil 2 a plus de pauses avant « A » et « C » supérieures à 2 s, alors que dans le profil 1 cette part est beaucoup plus faible, et équivalente pour les deux types de pauses. Globalement, on voit nettement que, pour tous les types de pauses, les pauses sont plus longues dans le profil 2 que dans le profil 1. En outre, la « hiérarchie » des emplacements syntaxiques (« P » > « A » > « C » > « I ») est reflétée différemment dans les deux profils. Dans le profil 1, la distinction entre « P » et « A » est très forte dans les catégories « 0-500 ms » et « 2000 ms-plus », alors qu'elle est moins marquée entre » A » et « C » dans les catégories « 0-500 ms » et inexistante dans la catégorie «2000 ms-plus ». La hiérarchie des emplacements syntaxiques est marquée de manière beaucoup plus linéaire dans le profil 2. Dans l'ensemble, l'hypothèse 1b est donc vérifiée.

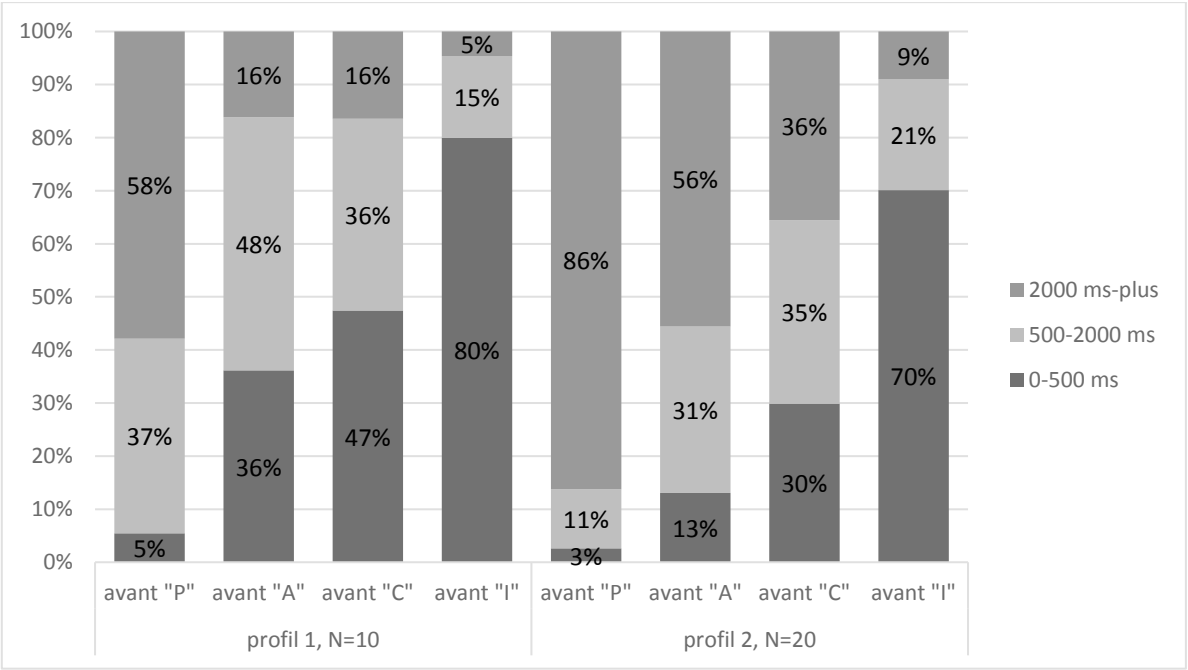


Figure 32. Proportion de pauses par grande catégorie de durée en fonction de l'emplacement syntaxique (durées en ms, moyennes en %)

1.2.3. Répartition du temps d'écriture

1.2.3.1. Description générale

De manière générale, plus de 60 % du temps d'écriture[58] est consacré aux pauses entre les mots ; le reste correspond à l'écriture des mots, comprenant les temps d'exécution motrice et les pauses intra-mots. Ces moyennes sont indiquées dans le Tableau 24 et illustrées par la Figure 33.

	pauses initiales	pauses intra-mot	exécution
répartition sur l'ensemble du texte	62% (11%)	23% (7%)	15% (6%)

Tableau 24. Répartition générale du temps d'écriture (moyennes en % et écart-types)

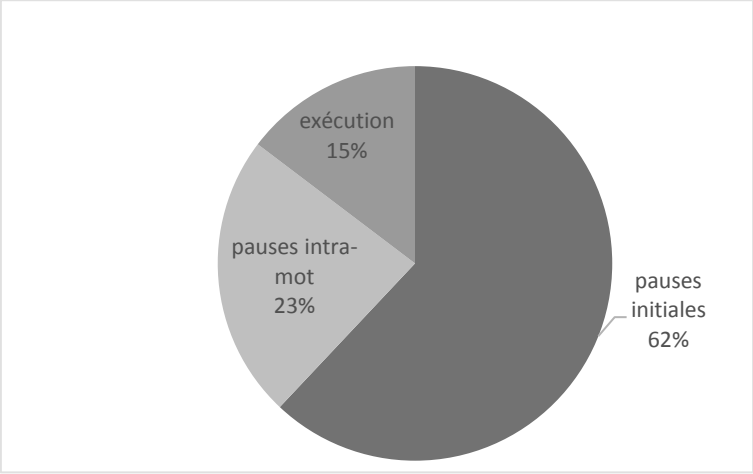


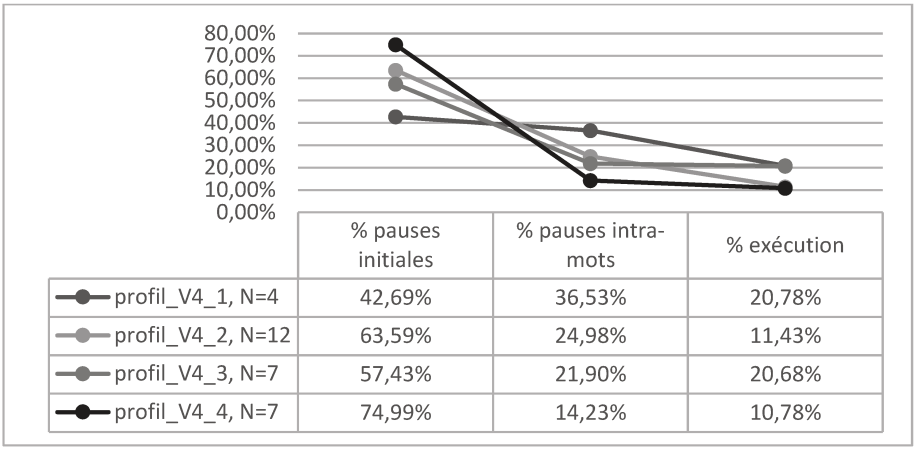
Figure 33. Répartition du temps d'écriture (moyennes en %)

Il y a une forte corrélation négative entre la part d'écriture dévolue aux pauses initiales et celle dévolue aux pauses intra-mot (r=-0,87) et à l'exécution motrice (r=-0,75). L'hypothèse 2a semble ainsi confirmée. En revanche, il n'apparait pas que les scripteurs faisant les pauses intra-mots les plus courtes aient un temps d'exécution plus lent : l'hypothèse 2b n'est donc pas confirmée.

1.2.3.2. Profils

Quatre profils ressortent de la classification hiérarchique. L'analyse discriminante classe à 100% les profils à partir des trois variables entrées, ces trois variables étant significativement différentes entre les groupes. Il apparait que deux variables permettent notamment de caractériser ces profils, comme résumé dans le Tableau 25 : la part dévolue aux pauses initiales (profil 4 > profils 2 et 3 > profil 1) et au temps d'exécution (profils 1 et 3 > profils 2 et 4).

[58] Le terme « temps d'écriture » inclut le temps mis pour produire le texte : il englobe donc les temps de pauses initiales, pauses intra-mots et exécution.



	% pauses initiales	% pauses intra-mots	% exécution
profil_V4_1, N=4	42,69%	36,53%	20,78%
profil_V4_2, N=12	63,59%	24,98%	11,43%
profil_V4_3, N=7	57,43%	21,90%	20,68%
profil_V4_4, N=7	74,99%	14,23%	10,78%

Figure 34. Profils liés à la répartition du temps d'écriture (moyennes en %)

V4	Répartition temps d'écriture	Profil_V4_1	N=4	40% pauses initiales, 20% exécution
		Profil_V4_2	N=12	65% pauses initiales, 10% exécution
		Profil_V4_3	N=9	60% pauses initiales, 20% exécution
		Profil_V4_4	N=7	75% pauses initiales, 10% exécution

Tableau 25. Répartition du temps d'écriture – Caractéristiques des profils (V4)

1.3. Évolution de la dynamique d'écriture au long du texte

Chaque texte a été segmenté en cinq parties de durée équivalente, une fois la pause de pré-écriture enlevée. Ainsi, il est possible de suivre l'évolution de l'écriture au fur et à mesure de l'accomplissement de la tâche. La « longueur » d'une partie désigne donc un nombre (ou pourcentage) de séquences, toutes les parties faisant la même durée.

1.3.1. Longueur de chaque partie

1.3.1.1. Description générale

De manière générale, les parties sont de longueur relativement équivalentes (Tableau 26) ; les parties 1 et 2 sont légèrement plus courtes (19% des séquences) que les parties suivantes, ce qui signifie que les scripteurs prennent légèrement plus de temps pour écrire durant ces deux parties.

		Proportion de séquences du texte par partie				
durée moyenne (en ms) d'écriture par partie	nombre moyen de séquences par partie	Partie 1	Partie 2	Partie 3	Partie 4	Partie 5
194814,93= 3 min et 14 sec (61782,28)	65,71 (16,83)	19% (3%)	19% (2%)	21% (3%)	20% (3%)	21% (4%)

Tableau 26. Longueur des parties (moyennes et écart-types)

Il n'y a pas de forte corrélation (r>0,7) entre ces variables, mais nous observons tout de même quelques corrélations significatives. Ainsi, il y a une corrélation négative entre la longueur de la partie 1 et de la partie 4 (r=0,58) et entre la longueur de la partie 3 et de la partie 5 (r=-0,69). En outre, les textes les plus longs, que ce soit en termes de durée d'écriture ou de nombre de séquences, sont davantage associés à des parties 3 courtes et à des parties 5 longues.

1.3.1.2. Profils

L'arbre issu de la classification hiérarchique fait ressortir 3 groupes. L'analyse discriminante permet de confirmer ces groupes, à part pour un sujet. Des différences significatives existent pour chacune des 5 parties. Globalement, les 5 parties sont de longueurs à peu près équivalentes : elles contiennent entre 16 % et 23% du texte. Le Tableau 27 résume les caractéristiques de chacun des profils, illustrés par la Figure 35.

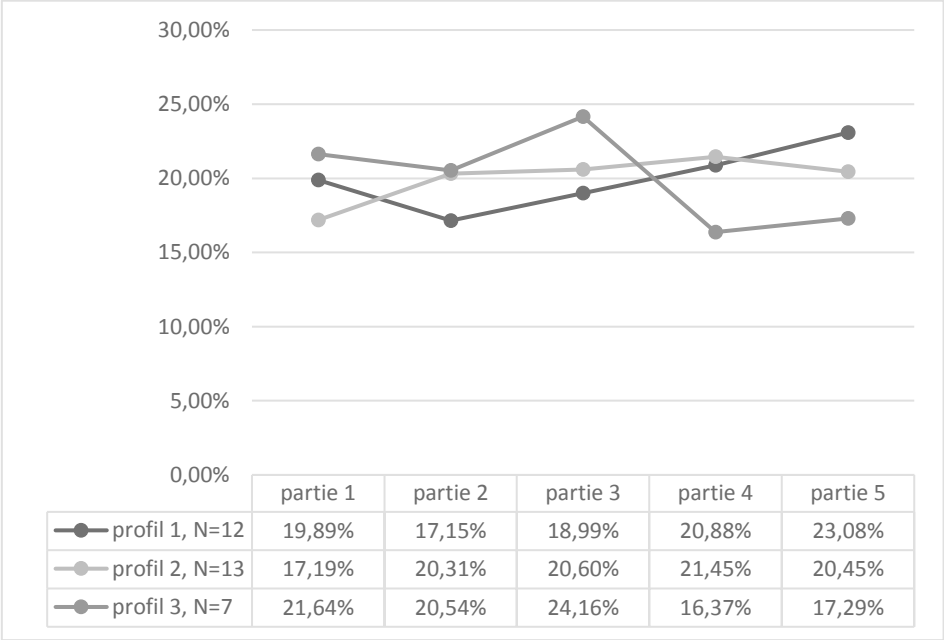


Figure 35. Profils liés à la proportion de texte écrit dans chaque partie (moyennes en %)

V5	Longueur des parties	Profil_V5_1	N=10	Partie 2 courte, puis augmentation, partie 5 longue
		Profil_V5_2	N=13	Partie 1 courte, puis stable
		Profil_V6_3	N=7	Irrégulier : partie 3 longue, parties 4 et 5 courtes

Tableau 27. Longueur des parties - Caractéristique des profils (V5)

1.3.2. Durées des pauses dans chaque partie

1.3.2.1. Description générale

Afin de savoir si les scripteurs faisaient des pauses plus ou moins longues suivant les parties de leur texte, nous avons calculé la différence entre la durée moyenne des pauses de la partie et la durée moyenne des pauses du texte (Tableau 28, Figure 36). La description générale montre peu de grandes

variations d'une partie à l'autre (au maximum, il y a 140 ms de différence entre la moyenne de la partie et la moyenne générale).

Pauses - Partie 1	Pauses - Partie 2	Pauses - Partie 3	Pauses - Partie 4	Pauses - Partie 5
+ 121,41 (153,62)	+ 140,15 (110,38)	-23,30 (-19,50)	+ 37,93 (-111,53)	+ 64,09 (-129,06)

Tableau 28 Différences entre la moyenne des pauses de chaque partie et la moyenne générale des pauses du texte (moyennes en ms et écart-types)

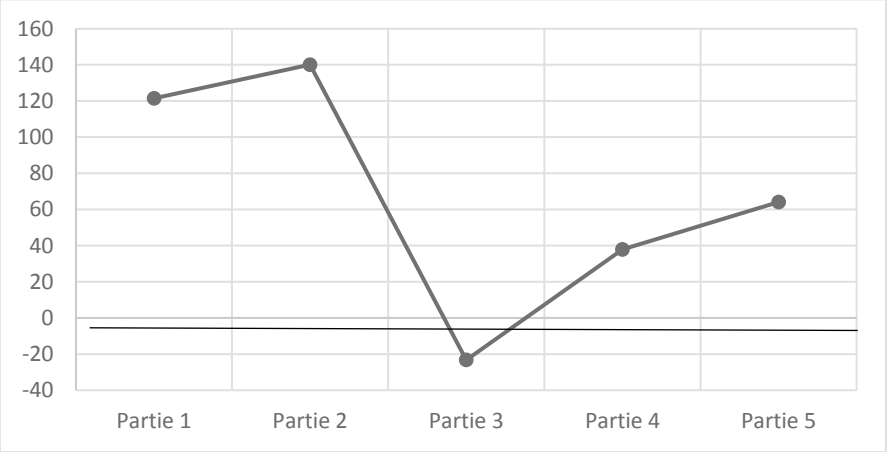
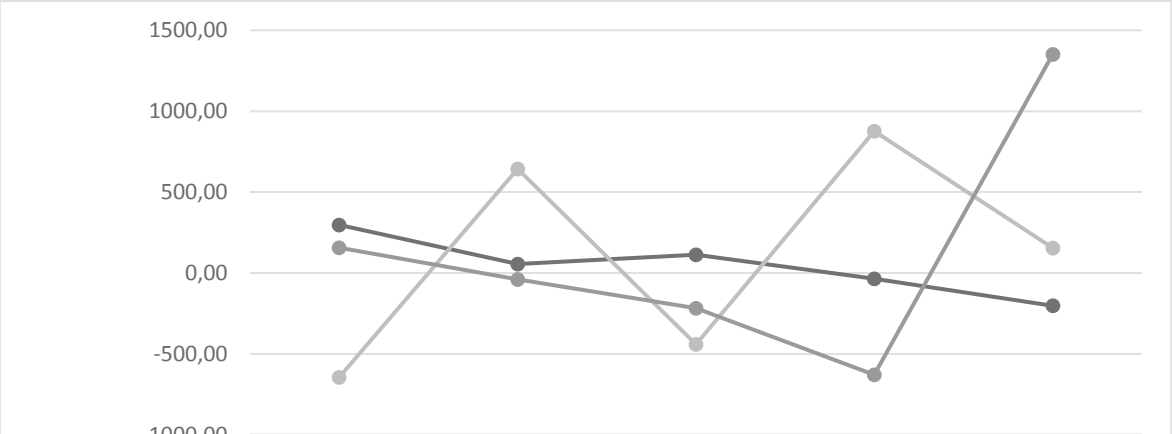


Figure 36. Différences entre la moyenne des pauses de chaque partie et la moyenne générale des pauses du texte (moyennes en ms)

On observe quelques corrélations significatives, toutes négatives : entre les différences de durée des pauses de la partie 1 et de la partie 4 (r=-0,65), et entre celles de la partie 5 et des parties 2 (r=-0,38) et 4 (r=-0,53).

1.3.2.2. Profils

Suite à la classification hiérarchique, une répartition en 3 profils a été effectuée. La très grande part des participants appartient au profil 1, qui correspond à ce qui a été observé dans la moyenne générale : peu de variations en moyenne par rapport à la moyenne du texte (moins de 200 ms de variation), avec globalement une légère diminution de la durée des pauses. Le profil 2 se caractérise par des moyennes de pauses très irrégulières : des pauses plus courtes que la moyenne dans les parties 1 et 3 (respectivement -645 et -442 ms) et des pauses plus longues dans les parties 2 et 4 (+ 643 et +877 ms). Enfin, les scripteurs du profil 3 font des pauses plus courtes surtout dans la partie 4 (-630 ms), mais se caractérisent essentiellement par des pauses très longues dans la partie 5 (+ 1352 ms par rapport à la moyenne).



	pauses - partie 1	pauses - partie 2	pauses - partie 3	pauses - partie 4	pauses - partie 5
profil 1, N=21	297,29	54,62	113,43	-34,57	-202,71
profil 2, N=5	-645,00	642,80	-442,00	876,60	153,80
profil 3, N=4	156,00	-39,50	-217,25	-629,25	1352,50

Figure 37. Profils liés à la différence entre la moyenne des pauses de chaque partie et la moyenne générale des pauses du texte (moyennes en ms)

V6	Durée des pauses par parties	Profil_V6_1	N=21	Stable, légère diminution
		Profil_V6_2	N=5	Pauses courtes dans les parties 1 et 3, pauses longues dans les parties 2 et 4
		Profil_V6_3	N=4	Pauses plus courtes dans les parties 3 et 4, pauses très longues dans la partie 5

Tableau 29. Durée des pauses par partie - Caractéristiques des profils (V6)

1.3.3. Répartition du temps d'exécution et de pause dans chaque partie

1.3.3.1. Description générale

L'observation générale de la répartition du temps d'écriture ne fait pas ressortir de différence importante d'une partie à l'autre, comme le montre le Tableau 30.

	pauses initiales		pauses intra-mot		exécution	
Partie 1	62%	(12%)	24%	(8%)	14%	(6%)
Partie 2	63%	(13%)	23%	(10%)	14%	(6%)
Partie 3	63%	(14%)	22%	(9%)	14%	(8%)
Partie 4	62%	(11%)	23%	(8%)	15%	(6%)
Partie 5	62%	(14%)	23%	(9%)	15%	(8%)

Tableau 30. Répartition du temps d'écriture par partie (moyennes en % et écart-types)

Il y a de très nombreuses corrélations significatives entre toutes ces variables. Tout d'abord, la proportion de pauses initiales dans chaque partie est corrélée avec celle des autres parties, et plus particulièrement celles des parties 1 avec 3 et 5 (r=0,74 et r=0,70), et celles des parties 2 avec 3

(r=0,72). De même, la proportion de pauses intra-mots dans chaque partie est corrélée avec celle des autres parties (0,46<r<0,67), tout comme la proportion d'exécution (0,60<r<0,84). Par ailleurs, il y a des corrélations négatives entre la proportion de pauses initiales et la proportion de pauses intra-mots et d'exécution, cette corrélation étant toujours supérieure à 0,70 au sein d'une même partie. En revanche, il n'y a pas de corrélation significative entre la proportion de pauses intra-mots et d'exécution, hormis entre les pauses intra-mots de la partie 1 et l'exécution des parties 1, 2 et 5 (0,39<r<0,55).

1.3.3.2. Profils

Afin d'établir des profils, nous avons regardé dans quelle mesure le pourcentage d'exécution variait d'une partie à l'autre. En effet, la section précédente nous a déjà permis d'étudier les variations de la durée moyenne des pauses ; cela nous permet à présent de nous intéresser davantage aux temps d'exécution.

L'arbre hiérarchique fait clairement ressortir deux profils, qui reflètent la distinction suivant la proportion globale de la durée de l'exécution, comme l'illustre la Figure 38 : les 2/3 des scripteurs (N=19) consacrent environ 20% de leur temps à l'exécution, alors que les autres en consacrent environ 10%. Les participants du profil 1 sont stables tout au long du texte quant au temps dévolu à l'exécution. En revanche, les participants du profil 2 consacrent une part plus importante à l'exécution dans les parties 4 et 5.

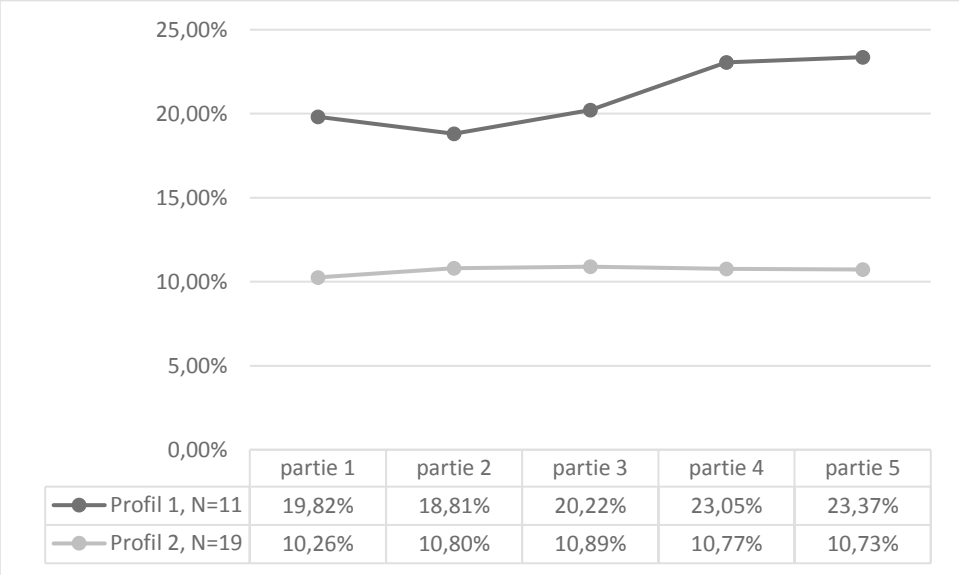


Figure 38. Profils liés à la proportion d'écriture consacrée à l'exécution par partie (moyennes en %)

Les caractéristiques de ces profils sont donc explicitées dans le Tableau 31.

V7	Proportion d'exécution par partie	Profil_V7_1	N=11	20 % exécution, parties 4 et 5 plus élevées
		Profil_V7_2	N=19	10% exécution, stable

Tableau 31. Proportion d'écriture consacrée à l'exécution par partie - Caractéristiques des profils (V7)

1.4. Résumé des profils

1.4.1. Variables retenues pour chaque scripteur

La description précédente a permis d'explorer un certain nombre de caractéristiques du comportement chronométrique des scripteurs de notre corpus. Suite à cela, certaines variables ont pu être regroupées ; d'autres se sont révélées peu informatives. Nous comptons donc 12 variables, rappelées dans le Tableau 32. Dans la dernière colonne de ce tableau nous indiquons les abréviations de la description des profils, utilisées pour la suite des analyses.

V1	Ponctuation	Profil_V1_1	N= 12	Part de pause faible avant les ponctuations précédant « P » et « I »	peu avant P et I
		Profil_V1_2	N= 12	Part de pause moyenne avant les ponctuations précédant « P » et « I »	moy avant P et I
		Profil_V1_3	N= 5	Part de pause faible avant les ponctuations précédant « P », part de pause très important avant les ponctuations précédant « I »	faible avant P, élevé avant I
V2	Durée pauses et exécution	Profi_V2_1	N= 11	Pauses initiales courtes, pauses intra-mots courtes et temps d'exécution élevé	b,b,h
		Profil_V2_2	N= 12	Pauses initiales moyennes, pauses intra-mots longues et temps d'exécution faible	m,h,b
		Profi_V2_3	N= 7	Pauses initiales longues, pauses intra-mots moyennes et temps d'exécution élevé	h,m,h
V3	Pauses suivant emplacement	Profil_V3_1	N= 10	pauses avant « I » plutôt entre 100 et 250 ms ; pauses avant « A » et « C » entre 250 et 1000 ms ; pauses avant « P » entre 2 et 10 s	bas_moy_haut
		Profil_V3_2	N= 20	Pauses plus longues	haut+
V4	Répartition temps d'écriture	Profil_V4_1	N= 4	40% pauses initiales, 20% exécution	40%,40%,20%
		Profil_V4_2	N= 12	65% pauses initiales, 10 % exécution	65%,25%,10%

		Profil_V4_3	N=9	60% pauses initiales, 20% exécution	60%,20%,20%
		Profil_V4_4	N=7	75% pauses initiales, 10 % exécution	75%,15%,10%
V5	Longueur des parties	Profil_V5_1	N=10	Partie 2 courte, puis augmentation, partie 5 longue	2 bas, 5 haut
		Profil_V5_2	N=13	Partie 1 courte, puis stable	1 bas, stable
		Profil_V5_3	N=7	Irrégulier : partie 3 longue, parties 4 et 5 courtes	3 haut, 4-5 bas
V6	Durée pauses par parties	Profil_V6_1	N=21	Stable, légère diminution	=, légère diminution
		Profil_V6_2	N=5	Pauses courtes dans parties 1 et 3, pauses longues dans parties 2 et 4	1 et 3 courtes, 2 et 4 longues
		Profil_V6_3	N=4	Pauses + courtes dans les parties 3 et 4, pauses très longues dans la partie 5	3 et 4 courtes, 5 très longue
V7	Proportion exécution par partie	Profil_V7_1	N=11	20 % exécution, parties 4 et 5 plus élevées	20%, 4 et 5 +
		Profil_V7_2	N=19	10% exécution, stable	10%, stable
V8	Durée pause de pré-écriture	en ms			
V9	Temps total d'écriture	en ms			
V10	Temps d'écriture sans pause de pré-écriture	en ms			
V11	Nombre total de séquences				
V12	Nombre total de révisions				

Tableau 32. Résumé des variables chronométriques

1.4.2. Lien entre les profils

Après avoir identifié des profils liés aux 7 groupes de variables (V1 à V7), le but est de voir si des profils plus généraux existent : il s'agit donc de voir s'il y a un lien de dépendance entre ces différents profils. Les résultats sont mentionnés dans le Tableau 33. Dans le cas de variables nominales, le test à notre disposition est le test du χ^2. Nous avons donc effectué des tests de χ^2 entre chacune des variables nominales. Les cases notées « **s** » indiquent que $p<0,05$; « tend » signifie que $0,5<p<0,10$; sinon nous avons noté « ns ». Le fait que le résultat du χ^2 soit significatif signifie que certaines distributions

observées diffèrent significativement des distributions théoriques : un lien de dépendance peut donc être noté entre les deux variables. Cependant, ces résultats doivent être pris avec précaution : les résultats de ce test sont peu fiables si les effectifs théoriques sont inférieurs à 5. Or, cela est le cas à plusieurs reprises. Nous présentons donc les résultats car ils peuvent donner des premières indications générales, mais nous allons regarder plus en détail les profils des participants pour vérifier la cohérence de leurs profils.

Les variables V8 à V12 sont continues. Pour voir s'il y avait un lien entre ces variables et les profils, nous avons comparé la distribution de ces variables continues en fonction des différents profils pour chaque groupe de variables (Test de comparaison de Mann-Whitney s'il y avait 2 groupes, Test de Kruskal-Wallis s'il y avait 3 groupes). Les variables V8 à V12 ont été comparées entre elles grâce à des tests de corrélation.

Variables	Ponctuation	Durée pauses et exécution	Pauses suivant emplacement	Répartition temps d'écriture	Longueur des parties	Durée pauses par parties	Proportion exécution par partie	Pause de pré-écriture	Temps total d'écriture	Temps d'écriture sans pré-écriture	Nb total de séquences	Nb de révisions
Ponctuation	/	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Durée pauses et exécution	ns	/	s	s	ns	ns	s	ns	**s** (0,048)	Tend (0,56)	ns	ns
Pauses suivant emplacement	ns	s	/	s	ns	s	s	ns	ns	ns	s	s
Répartition temps d'écriture	ns	s	s	/	s	ns	s	ns	Tend (0,066)	Tend (0,073)	ns	ns
Longueur des parties	ns	ns	ns	s	/	s	ns	Tend (0,09)	ns	ns	s	ns
Durée pauses par parties	ns	ns	s	ns	s	/	ns	s	Tend (0,088)	ns	s	ns
Proportion exécution par partie	ns	s	s	s	ns	ns	/	ns	Tend (0,052)	Tend (0,052)	Tend (0,064)	ns
Pause de pré-écriture	ns	ns	ns	ns	Tend (0,09)	s	ns	/	s	s	ns	ns
Temps total d'écriture	ns	**s** (0,048)	ns	Tend (0,066)	ns	Tend (0,088)	Tend (0,052)	s	/	s	s	ns
Temps d'écriture sans pré-écriture	ns	Tend (0,056)	ns	Tend (0,073)	ns	ns	Tend (0,052)	s	s	/	s	ns
Nb total de séquences	ns	ns	s	ns	s	s	Tend (0,064)	ns	s	s	/	s
Nb de révisions	ns	ns	s	ns	ns	ns	ns	ns	ns	ns	s	/

Tableau 33. Liens entre les variables

Plusieurs remarques émergent de ces résultats. Nous allons à présent les regarder en détail, en essayant de dégager des ensembles cohérents.

Tout d'abord, les profils concernant la ponctuation ne sont reliés à aucun autre profil. Nous n'avions pas d'hypothèse particulière pour cette variable chronométrique car elle a été rarement explorée dans les études précédentes ; les résultats obtenus n'ouvrent pas de nouvelles perspectives à explorer, du moins chez les adultes.

Il y a ensuite un ensemble de variables liées entre elles, concernant la durée des pauses et des temps d'exécution, ainsi que la répartition du temps d'écriture : un lien de dépendance existe entre les variables générales « Durée pauses et exécution », « Pauses suivant emplacement », « Répartition temps écriture », « Proportion exécution par partie ». Les différents cas de figure relevés (« types ») sont indiqués dans le Tableau 34. Les valeurs soulignées une fois indiquent des pauses initiales plutôt courtes, alors que le double soulignement indique des pauses initiales plutôt longues ; les valeurs en gras indiquent une part d'exécution importante.

Participants	Pauses suivant emplacement	Durée pauses et exécution	Répartition temps d'écriture	Proportion exécution par partie	Type
N=3 : S1, S20, S27	bas_moy_haut	b,b,**h**	40% ,40%, **20**%	20%	A1
N=2 : S10, S30	bas_moy_haut	m,h,b	65%, 25%, 10%	10%	A2
N=3 : S3, S4, S26	bas_moy_haut	b,b,**h**	60%, 20%, **20**%	20%	A3
N=1 : S7	bas_moy_haut	b,b,**h**	75%, 15%, 10%	20%	A4
N=1 : S17	bas_moy_haut	m,h,b	40%, 40%, **20**%	10%	A5
N=4 : S11, S13, S16, S19	haut+	b,b,**h**	60%, 20%, **20**%	20%	A6
N=9 : S2, S5, S12, S14, S15, S18, S23, S28, S29	haut+	m,h,b	65%, 25%, 10%	10%	A7
N=1 : S22	haut+	h,m,**h**	65%, 25%, 10%	10%	A8
N=6 : S6, S8, S9, S21, S24, S25	haut+	h,m,**h**	75%,15%,10%	10%	A9

Tableau 34. Lien entre les profils : moyenne des temps de pauses et exécution

Parmi ces types relevés, la plupart sont cohérents :

- Les durées d'exécution « basses » sont généralement associées à « 10% » d'exécution (types A2, A7) et les durées d'exécution « hautes » avec « 20% » (types A1, A3, A4, A6), sauf dans le cas où les pauses sont également très longues, ce qui fait que la proportion d'exécution par rapport à l'écriture totale reste faible (types A8, A9).

- Les pauses généralement plus longues sont associées à une proportion de pause initiale élevée (types A8, A9), alors que si elles sont moins longues, la proportion de pause initiale est plus faible ou moyenne (types A1, A2, A3, A5).

Les types A1, A2, A3, A7, A8, A9 sont donc tout à fait cohérents, ce qui représente 24 participants.

Les autres types, qui pourraient paraitre moins cohérents, peuvent s'expliquer :

- type A5 (S17) : la répartition du temps d'écriture est en réalité « 43%, 43%, 13% » : c'est pourquoi il est associé d'une part au profil « 40%, 40%, 20% », et d'autre part à 10% d'exécution par partie et un temps d'exécution plutôt bas.
- type A4 (S7) : ce sujet se caractérise par des pauses intra-mots faibles, ce qui fait qu'il est associé d'une part aux profils « b,b,h » et d'autre part au « 75%, 15%, 10% », ce qui parait contradictoire lorsqu'on regarde uniquement les pauses initiales et les temps d'exécution.
- type A6 (S11, S13, S16, S19,) : le profil 2 de la variable « pauses suivant emplacement » se caractérise certes par des pauses globalement plus longues, mais également par une proportion de « I 250-500 » élevée. Or cette proportion est très élevée chez les scripteurs de type 6, et leur durée d'exécution est également longue, ce qui explique l'association des profils « haut + » et « b, b, h ».

Il serait également cohérent que les autres variables liées entre elles concernent la longueur des parties et la gestion chronométrique au fur et à mesure du texte. Il apparait qu'il y a un lien de dépendance entre la longueur des parties et la façon dont les pauses sont plus ou moins longues par rapport à la moyenne. Un lien significatif a été relevé entre la longueur des parties et la répartition du temps d'écriture, mais pas avec la proportion d'exécution par partie. Le Tableau 35 indique les différents cas de figure répertoriés.

	Longueur des parties	**Durée pauses par parties**	**Proportion exécution par partie**	**Répartition temps d'écriture**	**Type**
N=2 : S1, S20	2 - , augm, 5 +	=, légère diminution	**20%**, 4 et 5 +	40%, 40%, **20%**	B1
N=1 : S7	2 - , augm, 5 +	=, légère diminution	**20%**, 4 et 5 +	75%, 15%, 10%	B2
N=5 : S12, S15, S18, S28, S30	2 - , augm, 5 +	=, légère diminution	10%, stable	65%, 25%, 10%	B3
N=1 : S9	2 - , augm, 5 +	=, légère diminution	10%, stable	75%, 15%, 10%	B4
N=3 : S5, S22, S29	2 - , augm, 5 +	1 et 3 courts, 2 et 4 longs	10%, stable	65%, 25%, 10%	B5
N=1 : S27	1 -,=	=, légère diminution	**20%**, 4 et 5 +	40%, 40%, **20%**	B6
N=5 : S3, S4, S11, S19, S26	1 -,=	=, légère diminution	**20%**, 4 et 5 +	60%, 20%, **20%**	B7
N=1 : S17	1 -,=	=, légère diminution	10%, stable	40%, 40%, 20%	B8
N=2 : S10, S23	1 -,=	=, légère diminution	10%, stable	65%, 25%, 10%	B9
N=1 : S21	1 -,=	=, légère diminution	10%, stable	75%, 15%, 10%	B10
N=1 : S13	1 -,=	3 et 4 court, 5 très long	**20%**, 4 et 5 +	60%, 20%, **20%**	B11
N=1 : S8	1 -,=	3 et 4 court, 5 très long	10%, stable	75%, 15%, 10%	B12
N=1 : S24	5 +, 4 et 5 -	=, légère diminution	10%, stable	75%, 15%, 10%	B13
N=2 : S6, 25	5 +, 4 et 5 -	1 et 3 courts, 2 et 4 longs	10%, stable	75%, 15%, 10%	B14
N=1 : S16	5 +, 4 et 5 -	3 et 4 court, 5 très long	**20%**, 4 et 5 +	60%, 20%, **20%**	B15
N=1 : S14	5 +, 4 et 5 -	3 et 4 court, 5 très long	10%, stable	65%, 25%, 10%	B16

Tableau 35. Lien entre les profils : comportement suivant les parties

L'appartenance à un certain profil « longueur des parties » peut généralement s'expliquer soit par la durée des pauses dans chaque partie, soit par la proportion de temps consacrée à l'exécution motrice : on peut supposer que plus les pauses sont courtes et/ou plus le temps d'exécution est long, plus il y aura de texte écrit dans la partie (double soulignement) ; à l'inverse, si les pauses sont plus longues, moins de texte sera écrit (simple soulignement).

Ainsi, les types B3 et B4 se caractérisent par une augmentation de la longueur des parties, et une partie 5 très longue : or, les pauses diminuent légèrement au fur et à mesure du texte. C'est également le cas des types B1 et B2, qui en outre ont une proportion d'exécution plus élevée dans la partie 5.

Le type B5 se distingue notamment par une partie 2 plus courte, ce qui est cohérent avec la présence de pauses longues dans cette partie. Les types B14, B15 et B16 ont une partie 3 longue, et des parties 4 et 5 courtes : on peut le relier au fait que les pauses des parties 3 soient plus courtes, et que les pauses des parties 4 (type B14) ou 5 (types B15 et B16) soient plus longues. Enfin, les types B8, B9, B10 correspondent à des textes dont la longueur des parties est à peu près équivalente tout au long du texte, avec une partie 1 plus courte ; or, le temps consacré à l'exécution est stable, ainsi que la durée des pauses, avec une légère diminution de celle-ci, qui peut expliquer que la partie 1 soit plus longue. De même, il est possible que la plus grande longueur de la partie 1 soit expliquée par la diminution des pauses au long du texte, et des parties 4 et 5 avec davantage d'exécution, dans le cas des types B6 et B7.

Concernant le lien entre la répartition d'écriture et les données liées aux parties, nous constatons tout d'abord que les scripteurs ayant un profil avec 20% d'exécution dans la répartition globale d'écriture l'ont aussi dans la proportion d'exécution par partie (hormis S7, dont le cas a été discuté précédemment). En outre, le profil de longueur « partie 2 courte, puis augmentation, partie 5 longue » est davantage lié au profil de répartition d'écriture « 65%, 25%, 10% » ; le profil « partie 1 courte, puis stable » est plus lié au profil « 65%, 20%, 20% », et le profil « partie 3 longue, parties 4 et 5 courtes » est lié au profil « 75%, 15%, 10% ».

Parmi les liens entre les profils, on remarque que la variable « pause suivant emplacement » est liée d'une part à la proportion d'exécution par partie, et d'autre part à la durée des pauses par partie. Ainsi, les pauses qui se répartissent de manière régulière (plus ou moins longues suivant l'emplacement : profil_V3_1) correspondent toutes au profil de pauses par partie « stable, légère diminution » (profil V6_1), alors que les profils plus irréguliers dans la longueur des pauses par parties (profil_V6_3) ont globalement des pauses plus longues (profil_V3_2). Cela confirme l'hypothèse 3a. Par ailleurs, le profil avec les pauses plus courtes va davantage correspondre au profil « 20% d'exécution, parties 4 et 5 plus longues » (profil_V7_1), alors que lorsque les pauses sont très longues, il s'agira plutôt du profil « 10% d'exécution, stable » (profil_V7_2). Ce profil de proportion d'exécution par partie est également directement relié à la variable « durée des pauses et d'exécution », puisque tous les profils « 20% d'exécution, parties 4 et 5 plus longues » correspondent au profil « b, b, h » (profil V2_1), alors qu'il n'a que 10% d'exécution dans le cas des profils « h,m,h » (profil V2_3) ou « m,h,b » (profil V2_2).

Le temps total d'écriture est corrélé positivement avec le temps d'écriture sans la pause de pré-écriture ($r=0,99$), ainsi qu'avec la pause de pré-écriture ($r=0,62$). Il est également lié au profil « durée des pauses et d'exécution » : la moyenne est plus élevée pour le profil « h,m,h » (profil V2_3) que

« b,b,h » (profil V2_2). D'autres tendances apparaissent : le temps total d'écriture est moins long pour le profil de répartition d'écriture « 40%,40%,20% » (profil_V4_1) que pour les autres, moins long pour les scripteurs du profil « pauses courtes dans parties 1 et 3, pauses longues dans partie 5 » (profil_V6_2), ainsi que dans le cas du profil « 20% d'exécution par partie » (profil_V7_1). Ces liens sont conservés lorsqu'on enlève la pause de pré-écriture du temps total d'écriture, hormis pour le lien avec la durée des pauses par partie.

Le temps de pré-écriture est justement lié à la variable « durée des pauses par partie » : le profil « pauses courtes dans parties 1 et 3, pauses longues dans parties 2 et 4 » (profil V6_2) correspond davantage à un temps de pré-écriture long.

Le nombre de séquences dans le texte est également lié à un certain nombre de variables. Cette variable est positivement corrélée au temps total d'écriture, avec ou sans pause de pré-écriture (respectivement r=0,55 et r=0,58). Elle est en outre liée à certaines variables concernant la durée des pauses. En effet, le texte est moins long avec le profil de durée suivant l'emplacement « pauses plus longues » (profil_V3_2), mais a tendance à être plus long lorsque la proportion d'exécution par partie est environ égale à 20% (profil_V7_1). Le texte est en outre plus court lorsque la durée des pauses correspond au profil « pauses + courtes dans les parties 3 et 4, pauses très longues dans la partie 5 » (profil_V6_3), ce qui est lié à un texte dont la longueur est de profil « partie 3 longue, parties 4 et 5 courtes » (profil_V5_3).

Enfin, le nombre de révisions dans le texte est positivement corrélé au nombre de séquences dans le texte (r=0,40). En outre, les scripteurs dont le profil concernant les pauses suivant l'emplacement correspond aux « pauses plus longues » (profil_V3_2) font globalement moins de révisions : de manière plus précise, le nombre de révisions est notamment corrélé positivement à « %P 1000-2000 » (r=0,48) et « %A 1000-2000 » (r=0,37), mais négativement corrélé aux pauses de durée plus longue, à savoir « % P 10000-plus » et « % A 2000-10000 ». Cela est conforme à l'hypothèse 3b.

1.5. Discussion sur la description générale des données chronométriques

Cette première étape dans notre étude expose de manière précise comment s'organise la gestion temporelle de la production écrite. Cela permet à la fois de caractériser l'écriture d'un corpus relativement homogène, puisque tous les scripteurs étaient des étudiants de langue maternelle française, sans trouble du langage, écrivant la suite d'un récit fictif, mais également de mettre en lumière l'existence de différences interindividuelles.

1.5.1. Ponctuation

Tout d'abord, concernant la ponctuation, il a été remarqué que la proportion de temps passé avant le signe de ponctuation différait suivant l'emplacement syntaxique, ce qui confirme l'hypothèse 1a. En effet, entre les phrases, les scripteurs s'arrêtent assez peu avant de mettre un signe de ponctuation. Pour ces scripteurs adultes, la phrase est vraisemblablement une unité de production : ils savent quand elle est achevée et peuvent la terminer assez rapidement en la ponctuant. Le temps long avant l'écriture du premier mot de la phrase peut avoir plusieurs fonctions dont celle souvent évoquée de planification (Matsuhashi, 1982), qui est généralement utilisée dans les études sur la temporalité de l'écriture. Cette pause peut parfois inclure un temps de relecture du texte déjà écrit (Fortier & Préfontaine, 1994) : ainsi, les retours sur le texte sont susceptibles d'être plus nombreux, plus éloignés et avec davantage de fixations en frontière de phrase qu'aux autres emplacement syntaxiques (Torrance et al., 2015). Si la phrase est relue dans un but de vérification, c'est sans doute en tant qu'un tout, pour vérifier des aspects aussi bien grammaticaux ou orthographiques que de plus « haut niveau », comme la cohérence. Foulin et al. (1989), qui avaient trouvé un résultat similaire, suggéraient également que le long temps de pause après l'écriture de la phrase sert à intégrer celle-ci en mémoire à long terme, conformément à la proposition de Daiute (1984, 1986).

En revanche, la proportion de temps passé avant le signe de ponctuation est plus longue avant les clauses ou unités A, puisque plus de la moitié du temps de pause total est avant le signe de ponctuation. Plusieurs explications peuvent être suggérées. Tout d'abord, les marques de démarcation des unités A ou des clauses sont moins claires qu'entre les phrases : le scripteur peut donc hésiter à inscrire un signe de ponctuation ou à utiliser un connecteur logique qui ne nécessite pas d'être précédé par une virgule. Il est aussi possible que des relectures à but de vérification soient plus systématiques en fin de clause : dans ce cas, si la relecture porte davantage au niveau de la formulation, il est possible qu'elle ait plutôt lieu avant de mettre le signe de ponctuation, qui marquerait la fin de production de l'unité.

Enfin, le temps de pause avant les ponctuations à l'intérieur des clauses est très variable, pouvant être en moyenne presqu'aussi court qu'avant les phrases ou très longs. Il est difficile de discuter de ces ponctuations intraclausales, car les cas sont très divers et non différenciés par notre codage : il est possible qu'une virgule précède un nom au cours d'une énumération et soit alors très vite marquée ; à l'inverse, avant un adjectif non restrictif, le scripteur peut mette plus de temps avant d'écrire une virgule.

Par ailleurs, les profils liés à la ponctuation n'ont aucun lien de dépendance avec les autres données de gestion temporelle du texte. Ces résultats nous incitent donc essentiellement à penser que les processus rédactionnels diffèrent fortement d'un emplacement syntaxique à l'autre, et font ressortir la phrase comme une unité de production considérée comme un « tout ».

1.5.2. Durée des pauses et des temps d'exécution

Plusieurs remarques peuvent ensuite être faites concernant la durée des pauses. Dans l'ensemble celles-ci sont très longues, la moyenne de celles avant les frontières d'unités syntaxiques (« C », « A », « P ») étant très souvent plus longue que 2 secondes, ce qui est le seuil choisi dans de nombreuses études (e.g. Olive & Cislaru, 2015; Wengelin, 2006). Les pauses avant les phrases sont globalement très longues : en moyenne 9 secondes, voire même une forte proportion de pauses au-delà de 10 secondes pour certains scripteurs. Torrance (2015) reprend une distinction entre deux types d'opérations : Cognitive Band operations et Rational Band operations, ces dernières se déroulant généralement durant les pauses qui durent plus de 10 secondes. Ces Rational Band operations correspondent à des opérations de macro-planification puisqu'il s'agit notamment de résolution de problème ou de réflexion générale. Cela signifierait que certains scripteurs s'engagent très souvent dans des processus de haut-niveau en début de phrase : les processus engagés diffèrent alors fortement de ceux en début d'autres unités, qui durent généralement moins de 10 secondes. Les pauses avant « I » ont une moyenne assez élevée, de presque 1 seconde. Cependant cette moyenne peut être causée par quelques valeurs extrêmes, et la grande part des pauses avant « I » (entre 70 et 80%) durent moins de 500 ms. D'ailleurs, notre division en catégories assez fines a fait ressortir une forte corrélation négative, visible également au niveau des deux profils qui ont émergé : certains scripteurs font des pauses avant « I » qui durent essentiellement entre 100 et 250 ms, alors que pour d'autres la durée est plutôt entre 250 et 500 ms. En considérant la dynamique d'écriture globale, il est possible de considérer que certaines pauses sont dues à l'activité graphomotrice, alors que d'autres sont des pauses cognitives (Chenu et al., 2014). L'émergence de ces deux types de comportement par rapport à la durée des pauses avant « I » peut alors avoir deux explications : soit certains scripteurs ont un seuil de pause « cognitive » plus élevé car leur dynamique d'écriture globale est plus lente, soit ils s'arrêtent délibérément plus longtemps en début de mot.

Par ailleurs, deux profils émergent nettement concernant la proportion de temps dévolue à l'exécution, et ce tout au long du texte : des scripteurs pour qui l'exécution motrice constitue 20% du temps d'écriture, et ceux pour qui cela constitue 10% du temps d'écriture. Le fait d'avoir considéré comme « pause » tout inactivité graphique supérieure à 15 ms peut expliquer que cette proportion de temps d'« exécution » paraisse faible au regard d'autres études qui indiquent par exemple que le scripteur passe 60% de son temps à écrire au clavier (Olive & Cislaru, 2015). Le fait que certains scripteurs passent plus de temps à l'exécution motrice que d'autres peut être lié à trois causes : d'une part il est possible que certains scripteurs fassent des pauses globalement courtes, d'où une proportion de temps d'exécution plus élevée ; d'autre part le temps d'exécution peut être effectivement plus long, soit car l'exécution motrice est plus difficile, soit au contraire parce qu'elle est relativement automatisée permettant l'activation d'autres processus cognitifs en parallèle, cela conduisant finalement à un ralentissement de cette exécution. Aucune de ces trois hypothèses ne peut être écartée, puisqu'il a été montré que des différences de compétences d'exécution existaient même chez les adultes (Connelly et

al., 2005), et que des processus de planification et de formulation pouvaient être à l'œuvre en même temps que l'exécution motrice (Olive et al., 2009). Nos données ne nous permettent pas de privilégier l'une des deux pistes ; une tâche de copie aurait pu permettre d'estimer les compétences motrices de nos scripteurs et de vérifier à quels types de scripteurs étaient associés les temps d'exécution les plus longs. Dans l'ensemble les profils des scripteurs étaient cohérents : les participants faisant les pauses initiales les plus longues avaient généralement une proportion d'exécution plus faible.

Dans notre corpus, les pauses de pré-écriture variaient en durée, mais duraient tout de même en moyenne plus d'une minute, ce qui peut laisser le temps d'une première planification générale. Le fait que la durée de cette pause soit positivement corrélée au temps total d'écriture peut effectivement suggérer que les principaux évènements du récit soient alors préparés.

1.5.3. Gestion temporelle en fonction de la longueur des textes

Une interrogation pouvait émerger quant à la gestion temporelle du texte en lien avec la longueur de celui-ci. Tout d'abord, les textes longs sont généralement associés à une proportion d'exécution élevée (environ 20%) et des pauses en fonction de l'emplacement syntaxique moins longues. Certes, il y a une corrélation positive entre le temps d'écriture total et le nombre de séquences par texte. Cependant, contrairement aux textes longs en nombre de séquences, un temps d'écriture élevé correspond généralement à des pauses initiales longues, et une proportion de temps d'exécution courte. Ces résultats pourraient paraitre contraires à ceux de Alves et al. (2007), qui montraient que les scripteurs les plus lents produisaient des textes plus courts et faisaient des pauses plus longues. Cependant, contrairement à ces auteurs qui parlent de « scripteurs lents » en se basant sur la médiane de l'intervalle de temps entre deux touches intra-mots, il s'agit dans notre cas des scripteurs dont la proportion de temps passé à écrire est plus élevée. Or, nous n'avons pas trouvé de corrélation significative directe entre la moyenne du temps d'exécution et la longueur des textes. Ces scripteurs qui écrivent des textes longs ne sont donc pas nécessairement les scripteurs les plus lents, mais ceux qui font le moins de pauses, comme c'était le cas pour des enfants avec ou sans trouble du langage (Connelly, Dockrell, Walter & Critten, 2012) et des enfants dyslexiques (Sumner, Connelly & Barnett, 2013). Cela est confirmé dans nos données par une corrélation négative entre la longueur du texte et la durée des pauses initiales, et notamment celles avant « I ».

En outre, il est apparu que les textes longs se caractérisaient par une fin de texte avec beaucoup de séquences, ce qui correspond à des pauses moins longues en fin de texte, et une partie centrale avec peu de séquences. Des études précédentes avaient également montré que les adultes expérimentés faisaient davantage de pauses en début qu'en fin de texte (Wengelin, 2006). Un pattern développemental était aussi apparu, avec les enfants de 9 à 12 ans écrivant beaucoup au début, les enfants de 15 ans étant stables, et les adultes écrivant beaucoup à la fois au début et à la fin de la session (Strömqvist, Holmqvist, Johansson, Karlsson & Wengelin, 2006). Cela pourrait s'expliquer par une étape de macro-planification

au milieu du texte dans le cas de textes longs, qui permettrait d'écrire la fin du récit plus rapidement. Une étude de Fortier et Préfontaine (1994) usant de protocoles verbaux rétrospectifs avait mis en évidence que les pauses dévolues à l'organisation conceptuelle et séquentielle étaient plus longues au milieu de la rédaction ; c'était aussi le moment où le scripteur ayant fait un plan au préalable le relisait le plus longtemps. Une autre hypothèse (n'excluant pas la première) serait que, le scripteur étant fatigué vers la fin d'un texte long, il accélère afin de finir plus rapidement.

1.5.4. Gestion temporelle en fonction de la partie du texte

Lorsque l'on regarde la dynamique d'écriture au fur et à mesure du texte, différents profils émergent : certains écrivent peu à la fin du texte, peut-être à cause de nombreuses relectures ; d'autres au contraire accélèrent et écrivent davantage vers la fin. Ces différences de rythme au long du texte s'expliquent à la fois par les variations de la durée des pauses, et par des variations dans l'exécution motrice. Cette dernière est cependant relativement stable : toujours autour soit de 10% soit de 20% du temps d'écriture, ceux dont l'exécution est la plus lente ayant tendance à ralentir encore plus vers la fin. Pour le comportement quant à la variation des durées des pauses, un des profils se caractérise par des grandes variations au cours du texte, comme si le texte était marqué à différents endroits par des très longues pauses. En effet, le profil correspondant aux fortes variations des moyennes des pauses en fonction des parties est lié à celui pour lequel les pauses en fonction des emplacements sont globalement très longues, et également à des pauses de pré-écriture en moyenne plus longues. Or, les pauses de pré-écriture sont positivement liées à la longueur des pauses avant « I » et avant « C ». Le fait de varier fortement la longueur des pauses n'est donc pas forcément dû à des pauses très longues avant les emplacements syntaxiques élevés, mais plutôt à des pauses longues, sans doute entre 2 et 10 secondes, en début de clauses.

1.5.5. Durée des pauses et révisions

Enfin, il a été remarqué que les textes comportant globalement des pauses très longues comptent moins de révision. Cela rejoint les résultats de Baaijen et a1. (2012), qui montraient également que plus les pauses en début de phrase étaient longues, moins il y avait de révisions : il est possible que certains scripteurs compensent un manque de planification initiale par des révisions plus nombreuses. Le fait que nos corpus comportent peu de révisions ne permet pas de se servir de celles-ci pour établir des profils clairs, comme cela a pu être fait dans des études précédentes (Hayes & Flower, 1980; van Waes & Schellens, 2003) ; nous reviendrons tout de même sur cet élément après avoir exploré le lien entre les pauses et la préparation des unités suivantes.

Cette première section nous a donc permis de nous familiariser avec la gestion temporelle globale du texte, et les caractéristiques quant à la durée des pauses et du temps d'exécution, en indiquant les

variations interindividuelles observées. Nous allons à présent nous intéresser plus particulièrement à la gestion temporelle de la micro-planification et de la formulation, donc à un niveau plus local. Pour cela, il est nécessaire de voir dans quelle mesure les unités de segmentation linguistique du texte peuvent être pertinentes pour étudier les processus d'écriture.

2. Les unités de segmentation du texte

Notre étude porte sur les « unités de production » du texte, et nous faisons l'hypothèse qu'il est possible de s'appuyer sur une segmentation basée sur des critères linguistiques pour mieux comprendre la gestion temporelle de l'écriture. Pour cela, nous verrons tout d'abord dans quelle mesure la durée des pauses varie en fonction de l'emplacement syntaxique : cela a été vérifié dans de nombreuses études ; notre segmentation est affinée avec la distinction entre « phrase » (critère graphique) et « unité A » (critère syntaxique). Nous avions également relevé que les liens de dépendance syntaxique pouvaient fournir des indications quant à l'empan de planification. Nous examinerons donc dans quelle mesure les diverses variables chronométriques (pauses initiales, pauses internes, temps d'exécution) varient en fonction de l'empaquetage syntaxique.

2.1. Longueur des pauses en fonction de l'emplacement syntaxique

La première étape concernant l'étude du lien entre unités de segmentation du texte et temporalité de l'écriture est de voir dans quelle mesure, dans nos données, les pauses les plus longues dans les textes correspondent à des frontières d'unités linguistiques. Pour cela, toutes les pauses inter-mots de chaque texte ont été prises en compte. Ces pauses ont été réparties d'une part en fonction de leur emplacement syntaxique (« P », « A », « C », « I »), et d'autre part en fonction de leur longueur relativement aux autres pauses du texte, en utilisant la répartition en déciles et en quartiles[59]. Après avoir décrit la répartition de ces pauses, nous avons voulu évaluer dans quelle mesure la longueur des pauses (en ms) pouvait être expliquée par leur emplacement syntaxique.

2.1.1. Répartition en quartiles

Le Tableau 36 indique le nombre de pauses en fonction de leur emplacement syntaxique et de leur durée catégorisée en quartile.

	Q1	Q2	Q3	Q4
P	5	1	23	**530**
A	5	13	53	**170**
C	31	55	**145**	**251**
I	**2181**	**2046**	**1945**	1223

Tableau 36. Répartition des pauses du texte en fonction de leur emplacement syntaxique et de la durée en quartile

[59] Qi désigne les pauses inférieures au quartile i : par exemple, « Q1 » correspond aux 25% de pauses les plus courtes.

Un test χ^2 a été effectué afin de vérifier si les variables « emplacement syntaxique » et « durée de la pause en quartile » sont indépendantes en probabilité. Le résultat du test est significatif ($\chi_{(9)}$=2282, p<0,001), ce qui signifie que ces variables sont dépendantes. Si ces deux variables étaient indépendantes, il devrait y avoir approximativement le même nombre de pauses par colonne au sein de chaque ligne. Nous avons indiqué en gras les cases pour lesquelles l'effectif observé est supérieur à l'effectif attendu : sur l'ensemble des données, il y a plus de « P » et de « A » qu'attendus en Q4, plus de « C » qu'attendu en Q3 et Q4, mais moins de « I » qu'attendu en Q4. Cela nous invite à considérer effectivement les frontières syntaxiques (« C », « A », « P) » comme des lieux privilégiés pour les pauses longues.

Parmi les 1223 pauses « I » qui font partie de Q4, 64 pauses correspondent à des débuts de révision, 57 pauses sont des reprises du texte après avoir effectué une révision, et 13 pauses correspondent à la dernière séquence du texte.

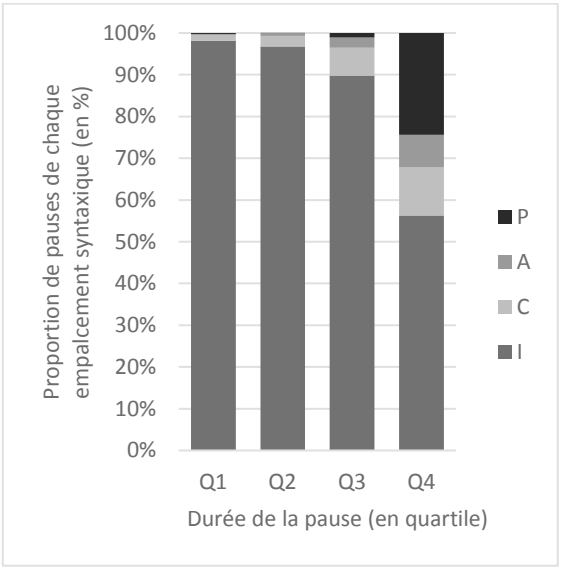


Figure 39. Proportion de pauses de chaque emplacement syntaxique par quartile (en %)

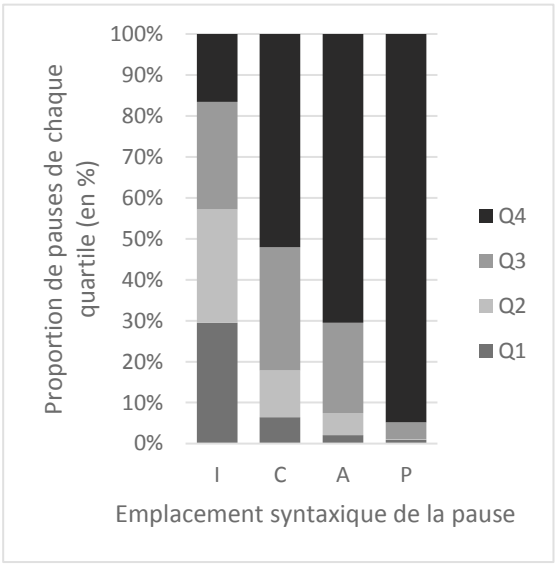


Figure 40. Proportion de pauses de chaque quartile par emplacement syntaxique (en %)

Les Figure 39 et Figure 40 illustrent bien que la quasi-totalité des pauses Q1 et Q2 correspond à des pauses « I », alors que près de 45% des pauses Q4 correspondent à des pauses « C », « A » ou « P ». En outre, bien que 55% des pauses Q4 correspondent à des pauses « I », cela ne représente que 18% de l'ensemble des pauses « I » ; à l'inverse, les 25% de pauses Q4 correspondant à des pauses « P » constituent 95% des pauses « P » des textes.

2.1.2. Répartition en déciles

L'observation de la répartition des pauses en quartiles a montré que les pauses « C », « A » et « P » étaient davantage présentes dans Q3 et Q4. La répartition en déciles permet d'affiner ces observations.

La Figure 41 représentant l'ensemble des données montre bien une diminution de la proportion de pauses « I » à partir de D6. En D10, plus de 40% des pauses correspondent à des pauses « P ».

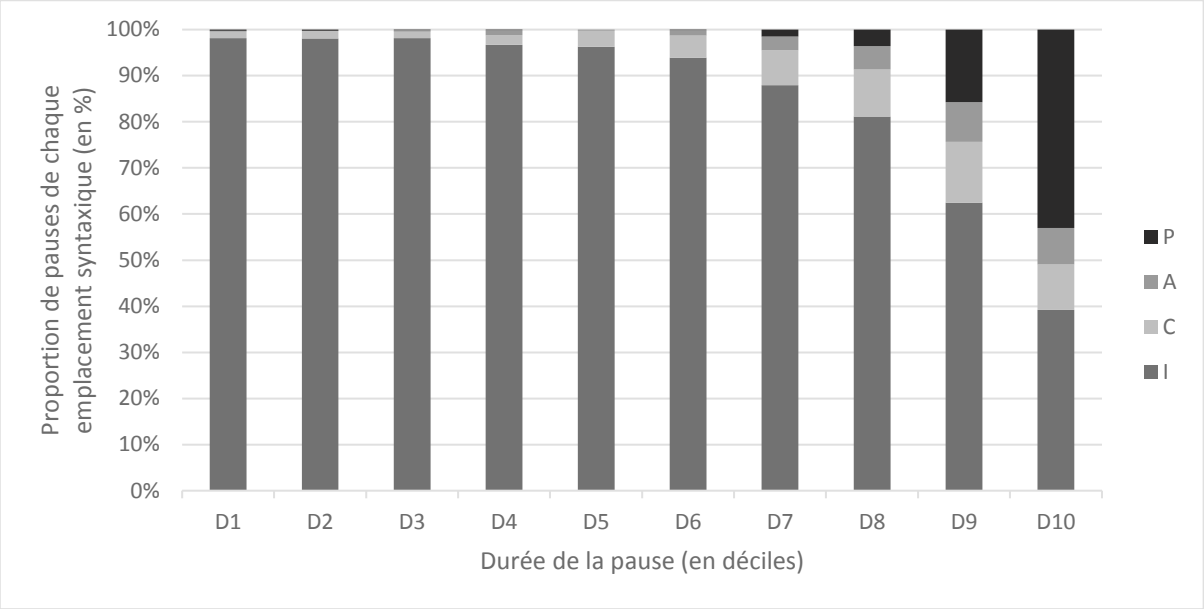


Figure 41. Proportion de pauses de chaque emplacement syntaxique par décile (en %)

Après avoir montré l'intérêt de distinguer les pauses « I » des pauses « C », « A », « P », nous pouvons nous interroger sur la pertinence de garder une distinction entre ces trois derniers types de pauses. Un premier élément de réponse peut être apporté en regardant uniquement la répartition de ces trois types de pauses parmi les pauses les plus longues, à savoir les déciles D8, D9, D10 (soit les 30% de pauses les plus longues).

Nous avons là encore effectué un test de χ^2, mais cette fois-ci en prenant uniquement en compte les trois frontières syntaxiques (« C », « A », « P ») et les trois déciles les plus longs. La question est alors de savoir si, parmi les pauses à des frontières syntaxiques faisant partie des pauses les plus longues du texte, il y a une dépendance entre le type de frontière syntaxique et la distribution dans les déciles D8, D9, D10. Le Tableau 37 indique les effectifs observés, et entre parenthèses les effectifs attendus.

	D8	D9	D10
« P »	31 (87)	137 (174)	**371** (278)
« A »	**44** (30)	**75** (60)	68 (97)
« C »	**88** (46)	**115** (93)	85 (149)

Tableau 37. Répartition des pauses en frontière syntaxique (« P », « A », « C ») les plus longues (D8 à D10)

Le résultat du test de χ^2 est significatif ($\chi_{(4)}$=163, p<0,001). Nous avons mis en gras les cases dans lesquelles l'effectif observé est supérieur à celui attendu : les pauses « P » apparaissent davantage dans D10, alors que les pauses « C » et « A » apparaissent davantage dans D8 et D9.

2.1.3. Longueur des pauses en fonction de l'emplacement syntaxique

Nous avons souhaité voir dans quelle mesure l'emplacement syntaxique expliquait la durée des pauses dans le texte. Pour cela, nous avons utilisé les modèles mixtes. Dans cette étude, l'effet fixe est l'emplacement syntaxique, qui est un facteur à plusieurs modalités. Les participants ont été entrés dans l'analyse en tant qu'effet aléatoire. La VD est la pause initiale.

2.1.3.1. Modèle avec quatre emplacements syntaxiques : « I », « C », « A », « P »

- Modèle 0 :

Le modèle vide nous donne un ICC de 0,0462, ce qui signifie que 4,62% de la variance peut être expliquée par les différences entre participants.

- Modèle 1 :

Dans le modèle, nous avons entré comme facteur fixe supplémentaire « l'emplacement syntaxique », à 4 modalités (« I », « C », « A », « P »). Ce modèle apporte une nette amélioration ($\chi_{(3)}$=3075,96, p< 0,001), chaque emplacement syntaxique étant significativement différent des autres. D'après ce modèle, 30,58% de la variance des pauses initiales peut être expliquée par l'emplacement syntaxique. La Figure 42 illustre les moyennes pour chacun de ces emplacements.

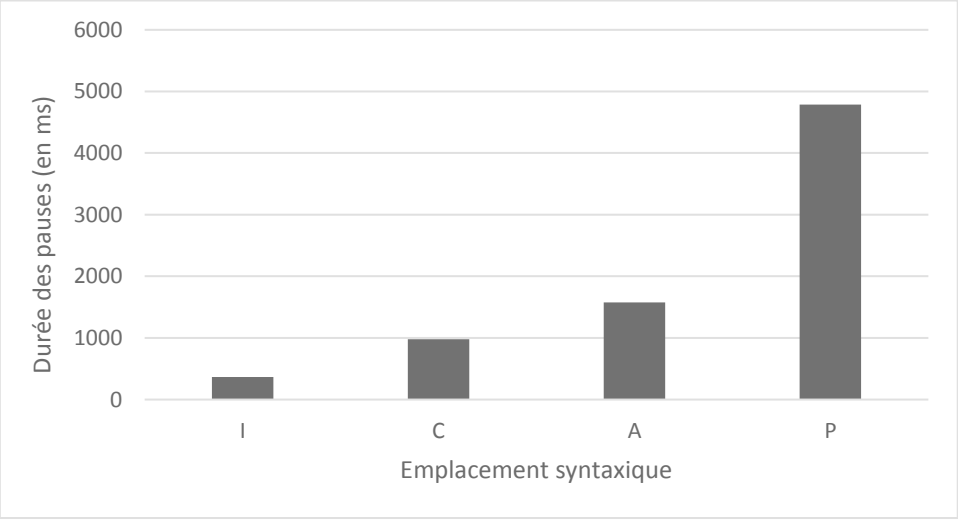


Figure 42. Durée des pauses en fonction de l'emplacement syntaxique (moyennes en ms)

2.1.3.2. Modèle avec trois emplacements syntaxiques : « C », « A », « P »

Les données introduites dans ce modèle sont uniquement les pauses situées en début de clause, que celle-ci soit en début de phrase « P », d'unité A (« A ») ou à l'intérieur d'une unité A (« C »).

- Modèle 0 :

Ce modèle donne un ICC (reflétant la part de variance expliquée par les différences interindividuelles) de 9,54%, ce qui justifie d'utiliser un modèle mixte, afin de prendre un compte cette variabilité due aux sujets.

- Modèle 1 :

Nous avons ensuite introduit les 3 emplacements syntaxiques « P », « A », « C ». Le modèle 1 est significativement meilleur que le modèle vide ($\chi_{(2)}$=380,49, p<0,001), chaque emplacement syntaxique étant significativement différent des autres. Dans ce modèle, le pseudo R^2 (reflétant la part de variance pouvant être expliquée par les effets fixes, donc ici l'emplacement syntaxique) est égal à 26,58%.

- Modèle 2 :

Afin de nous assurer de la pertinence de distinguer entre « A » et « C », nous avons fait un nouveau modèle dans lequel nous avons entré le facteur « emplacement syntaxique » avec seulement deux modalités : « P » et « A + C ». Ce modèle est significativement moins bon que le modèle 1 ($\chi_{(1)}$=26,59, p<0,001).

Le meilleur modèle est donc celui avec l'introduction des trois emplacements syntaxiques « P », « A » et « C ».

2.1.4. Discussion sur l'emplacement syntaxique

Ces premières analyses montrent donc que la longueur des pauses liées à des frontières syntaxiques n'est pas répartie au hasard. D'une part, les pauses « C », « A » et « P » sont davantage présentes dans les quartiles Q3 et Q4. D'autre part, au sein des pauses les plus longues, les pauses D10 semblent être privilégiées par les pauses interphrastiques, tandis que les pauses « C » et « A » se situent davantage dans D8 et D9. À ce stade, il semble donc pertinent de distinguer au moins 3 niveaux syntaxiques : les mots en début de phrase (« P »), les mots en début de clause qui ne sont pas des débuts de phrase (« A » et « C ») et les mots à l'intérieur des clauses (« I »).

Cependant, l'analyse de l'effet de l'emplacement syntaxique sur la longueur des pauses, effectuée grâce à des modèles linéaires mixtes, montre qu'il convient de maintenir la distinction entre « P », « A » et « C », puisque le modèle est moins explicatif si on ne conserve que deux niveaux de frontières syntaxiques (entre phrases et entre clauses à l'intérieur des phrases). Si l'on regarde à nouveau les quartiles, on voit effectivement que seuls 50% des pauses « C » sont des pauses Q4, alors que c'est le cas de 70% des pauses « A » : les pauses « A » sont en très grande majorité très longues (Q4) alors que les pauses « C » sont parfois bien moins longues (Q3 voire Q2). Ces résultats corroborent l'« effet syntaxique » révélé dans plusieurs études (e.g. Foulin, 1998; Schilperoord, 1996).

2.2. Longueur des pauses en fonction du lien syntaxique entre les clauses

Une explication souvent avancée pour justifier l'effet de l'emplacement syntaxique est la planification : il est plus long de planifier une phrase qu'une unité A et qu'une clause, d'où la différence de longueur de la pause initiale. Si c'est le cas, cela veut dire que la pause en début de phrase sert à planifier, du moins dans une certaine mesure, les clauses contenues dans cette phrase. On peut donc supposer, à la suite de van Hell et al. (2008) que les clauses ayant un lien de dépendance fort avec les clauses précédentes sont planifiées de manière anticipée. Cependant, il est possible que cette anticipation se déroule parfois en parallèle de l'écriture de l'unité précédente. Nous allons donc voir dans quelle mesure le lien de dépendance entre les clauses permet d'expliquer la variance des pauses initiales et du temps d'écriture.

2.2.1. Hypothèses

Si les clauses sont planifiées d'une part durant la pause initiale des grandes unités syntaxiques et d'autre part durant l'écriture de la clause précédente, deux hypothèses générales peuvent être testées :

- HG 1 : Plus il y a d'éléments à planifier, plus la pause initiale est longue, et plus le temps d'écriture est long.
- HG 2 : Plus le contenu a été planifié auparavant, plus la pause initiale est courte et plus le temps d'écriture est rapide.

Dans la première partie des analyses, les variables explicatives testées sont le nombre d'éléments dans l'unité (e.g. nombre de clauses dans l'unité A) et le numéro de l'unité par rapport à l'unité supérieure (e.g. numéro de l'unité A dans la phrase). À partir de là, deux séries d'hypothèses peuvent être posées :

- Hyp 1 : La longueur de la pause initiale varie en fonction du nombre d'éléments à planifier
    - Hyp 1a : La longueur de la pause initiale de la phrase est positivement corrélée avec le nombre d'unités A et / ou de clauses dans la phrase.
    - Hyp 1b : La longueur de la pause initiale de l'unité A est positivement corrélée avec le nombre de clauses dans l'unité A.
- Hyp 2 : La longueur de la pause initiale varie en fonction de sa position dans l'unité supérieure
    - Hyp 2a : La pause initiale de la 1ère unité A de la phrase est plus longue que celle des unités A suivantes. Cela équivaut à « la pause initiale de l'emplacement « P » est plus longue que celle de l'emplacement « A » ».

- Hyp 2b : La pause initiale de la 1ère clause de l'unité A est plus longue que celle des clauses suivantes. Cela équivaut à « la pause initiale de l'emplacement « A » est plus longue que celle de l'emplacement « C » ».

Dans la deuxième partie des analyses, les variables explicatives supplémentaires testées sont :

- le type syntaxique de la clause (juxtaposée, coordonnée, circonstancielle, relative, complétive, corrélative, circonstancielle avant matrice, juxtaposée après circonstancielle)
- clause matrice / non matrice
- profondeur syntaxique (indépendante / subordonnée)
- subordonnée enchâssée / non enchâssée
- subordonnée finie / non finie
- clause obligatoire ou non (facultative d'un point de vue communicatif /obligatoire d'un point de vue valenciel / référentiel / communicatif)

Les variables ci-dessus ont été explicitées dans le chapitre « Méthodologie générale », hormis pour l'aspect « obligatoire », qui sera explicité dans le chapitre suivant. Cette variable a permis de distinguer 4 types de clauses subordonnées en fonction de la nécessité de leur présence d'un point de vue syntaxique, référentiel ou communicatif : certaines clauses sont obligatoires pour compléter la valence du prédicat (OV) (55), pour délimiter le référent (OR) (56), nécessaires d'un point de vue discursif (OC) (57), alors que d'autres sont facultatives (FC) (58).

(55) Dimitri vit / **que les étoiles avaient disparu**. [S03]

(56) Sa main gauche était toujours bandée depuis la dernière altercation ***/ qu'il avait eue*** **avec Dimitri.** [S03]

(57) Dimitri assomma le gardien / **qui tomba évanoui**. [S15]

(58) Et ils partirent tous les deux en bateau pour la Crête / **où Dimitri possédait une ferme léguée par son parrain**. [S19]

Les analyses portent d'une part sur les pauses initiales de clause, et d'autre part sur les différentes composantes de l'écriture de la clause (moyenne des pauses inter-mots, moyenne des pauses intra-mots, moyenne du temps d'exécution, temps total d'écriture interne). Conformément aux hypothèses générales présentées ci-dessus, deux séries d'hypothèses peuvent être à nouveau testées sur les pauses initiales des clauses :

- HG 1 : Plus il y a d'éléments à planifier, plus la pause initiale est longue.

  - Hyp 1c : La pause initiale d'une clause matrice est plus longue que celle d'une clause non matrice.
  - Hyp 1d : La pause initiale d'une subordonnée avant matrice est plus longue que celle des autres subordonnées.
- HG 2 : Plus le contenu a été planifié auparavant, plus la pause initiale est courte.
  - Hyp 2c : La pause initiale d'une clause subordonnée est plus courte que celle d'une clause indépendante.
  - Hyp 2d : La pause initiale d'une subordonnée enchâssée est plus courte que celle d'une subordonnée non enchâssée.
  - Hyp 2e : La pause initiale d'une subordonnée non finie est plus courte que celle d'une subordonnée finie.
  - Hyp 2f : La pause initiale d'une subordonnée obligatoire est plus courte que celle d'une subordonnée facultative.

Par ailleurs, deux séries d'hypothèses peuvent être testées sur les variables chronométriques d'écriture de la clause :

- HG 1 : Plus il y a d'éléments à planifier, plus le temps d'écriture de la clause est long.
  - Hyp 1e : L'écriture d'une clause matrice est plus longue que celle d'une clause non matrice.
  - Hyp 1f : L'écriture d'une subordonnée avant matrice est plus longue que celle des autres subordonnées.
- HG 2 : Plus le contenu a été planifié auparavant, plus le temps d'écriture de la clause est court.
  - Hyp 2g : L'écriture de la 1ère unité A ou clause de la phrase est plus lente que celle des suivantes.
  - Hyp 2g : L'écriture d'une clause subordonnée est plus courte que celle d'une clause indépendante.
  - Hyp 2h : L'écriture d'une subordonnée enchâssée est plus courte que celle d'une subordonnée non enchâssée.
  - Hyp 2i : L'écriture d'une subordonnée non finie est plus courte que celle d'une subordonnée finie.
  - Hyp 2j : L'écriture d'une subordonnée obligatoire est plus courte que celle d'une subordonnée facultative.

Il n'est bien sûr pas attendu que toutes les hypothèses ci-dessus soient vérifiées : il est peu vraisemblable que les planifications de chaque type de clause affectent toutes les données chronométriques. Cependant, ces hypothèses permettent d'expliciter dans quels sens les résultats sont

attendus. L'intérêt de cette partie sera alors d'affiner la manière dont les liens entre les clauses affectent la gestion temporelle de l'écriture des unités linguistiques.

2.2.2. Pause initiale de la phrase

Nous avons regardé si la longueur de la pause initiale de la phrase variait en fonction du nombre de clauses et du nombre d'unités A dans la phrase. D'après le modèle vide, 18.94% de la variance est due aux différences inter-participants. Aucun des modèles testés avec les facteurs « nombre de clauses par phrase » et « nombre d'unités A par phrase » ne s'est révélé être significatif. L'hypothèse 1a n'est pas validée.

2.2.3. Pause initiale de l'unité A

Nous avons ensuite testé si la longueur de la pause en début d'unité dépendait, d'une part du nombre de clauses dans l'unité A, et d'autre part du numéro de l'unité A dans la phrase. Pour cela, nous avons entré comme données toutes les pauses en début d'unité A, hormis lorsqu'il n'y avait qu'une seule unité A dans la phrase.

- Modèle 0 :

Le modèle vide indique un ICC de 16,59%.

- Modèle 1 :

Le meilleur modèle obtenu à partir des facteurs « numéro de l'unité A dans la phrase » et « nombre de clauses dans la phrase » est celui où est entré seulement le facteur concernant la position. Ce modèle est significativement meilleur que le modèle vide ($\chi_{(3)}$=44,439, p<0,001), ce facteur apportant un gain explicatif de 29,98%. Les tests post-hoc suggèrent que la pause avant l'unité A qui est numéro 1 de la phrase est plus longue que les autres, sans que la différence soit significative entre « unité A numéro 1 » et « unité A numéro 4 ».

- Modèle 2 :

Afin d'affiner le modèle, nous avons donc testé le facteur « numéro de l'unité A dans la phrase » à deux modalités (1 vs 2 ,3, 4) et à trois modalités (1 vs 2, 3 vs 4). Le meilleur modèle est celui à deux modalités. En effet, il n'y a pas de différence significative entre ce modèle 2 et le modèle 1, mais le modèle 2 a l'avantage d'avoir 2 paramètres en moins. Par rapport au modèle 0 la différence est toujours significative ($\chi_{(1)}$=42,99, p<0,001), avec pseudo R^2=28,90%.

Ce modèle réplique donc simplement la différence de longueur entre pause « P » (unité A en début de phrase) et pause « A » (unité A à l'intérieur d'une phrase). L'hypothèse 1b n'est pas validée, mais l'hypothèse 2a est validée.

2.2.4. Pause initiale de la clause

2.2.4.1. En fonction de la place dans la phrase et dans l'unité A

Nous avons voulu vérifier l'effet du numéro de la clause dans l'unité A et dans la phrase sur la longueur de la pause initiale. Les données entrées dans le modèle sont toutes les pauses en début de clause, hormis lorsqu'il n'y avait qu'une seule clause par phrase.

- Modèle 0 :

Le modèle vide indique un ICC de 10,26%.

- Modèle 1 :

Le meilleur modèle obtenu à partir des facteurs « numéro de la clause dans la phrase » et « numéro de la clause dans l'unité A » est celui où les effets principaux de ces deux facteurs sont entrés simultanément. Ce modèle est significativement meilleur que le modèle vide ($\chi_{(12)}$=361,66, p<0,001), ces facteurs expliquant 30,04% de la variance totale. Les tests post-hoc suggèrent que la pause en début de phrase ou d'unité est plus longue que les suivantes, bien que la différence ne soit pas significative entre la clause « numéro 1 » et les clauses plus éloignées.

- Modèle 2 :

Afin d'affiner le modèle, nous avons donc testé le facteur « numéro de la clause dans la phrase » à deux modalités (1 vs 2 à 7) et à trois modalités (1 vs 2, 3, 4, 5 vs 6, 7), puis le facteur « numéro de la clause dans la phrase » à deux modalités (1 vs 2 à 7) et à trois modalités (1 vs 2, 3, 4 vs 5, 6, 7). Dans les deux cas, le facteur à deux modalités n'est pas significativement moins bon que celui à 3 ou à 7 modalités, et il présente l'avantage d'avoir moins de paramètres. Au final, le meilleur modèle est celui où sont introduits simultanément les effets principaux des deux facteurs à deux modalités : « numéro de la clause dans la phrase _ 2 » et « numéro de la clause dans l'unité A _ 2». En effet, il n'y a pas de différence significative entre ce modèle 2 et le modèle 1, mais le modèle 2 a l'avantage d'avoir 10 paramètres en moins. Par rapport au modèle 0, la différence est toujours significative ($\chi_{(2)}$=345,51, p<0,001), avec pseudo R^2=28,97%.

Là encore, ce modèle corrobore les résultats précédents, qui indiquaient une différence de longueur entre pauses « P » (clauses en début de phrase), pauses « A » (clauses en début d'unité A) et pauses « C » (pauses à l'intérieur d'unité A). L'hypothèse 2b est vérifiée.

2.2.4.2. En fonction d'autres critères syntaxiques

Outre l'emplacement syntaxique, d'autres critères peuvent affecter la durée des données chronométriques, comme le fait que la clause soit matrice ou non, son type syntaxique, etc. C'est ce que nous avons voulu tester dans les modèles suivants.

- Modèle 1 :

Nous avons entré simultanément les facteurs « numéro de la clause dans la phrase _ 2 », « numéro de la clause dans l'unité A _ 2 », « matrice », « obligatoire », « type syntaxique », « profondeur syntaxique », « non finie ». Les facteurs ayant un effet significatif sont : « numéro de la clause dans la phrase _ 2 », « matrice », « obligatoire », « type syntaxique ». Le modèle 1 dans lequel sont entrés les effets principaux de ces facteurs est significativement meilleur que le modèle vide ($\chi_{(13)}$=426,46, p<0,001), avec pseudo R²=30,49%. Les tests post-hoc suggèrent que l'on pourrait regrouper les types syntaxiques « relative », « complétive » et « circonstancielle (après matrice) » dans une même catégorie « subordonnée ». Par ailleurs, les modalités « obligatoire d'un point de vue valenciel » et « obligatoire d'un point de vue référentiel » pourraient être regroupées au sein d'une même catégorie.

- Modèle 2 :

Ce modèle teste l'entrée des mêmes facteurs, avec les catégories telles qu'elles ont été regroupées. Il est aussi bon que le modèle 1, mais avec 3 paramètres en moins. Il est également significativement meilleur que le modèle 0 (χ(10)=424,796, p<0,001).

Si on essaie d'enlever certains de ces facteurs, le modèle est toujours moins bon. Il s'agit donc du modèle le plus explicatif et le plus parcimonieux.

Ce modèle permet donc de valider les hypothèses 1c, 1d, 2b et 2f.

Le Tableau 38 ci-dessous indique les effets de chaque facteur au sein du modèle. Les graphiques qui suivent illustrent les différences significatives entre les différentes modalités au sein de chaque facteur. Comme attendu, les clauses en début de phrase ont une pause initiale plus longue (Figure 43). Les clauses matrices sont précédées d'une pause plus longue que les non-matrices (Figure 44), et les clauses « obligatoires syntaxiquement » (OV et OR) sont précédées d'une pause plus courte que celles facultatives d'un point de vue communicatif (Figure 45). Enfin, les clauses « circonstancielle avant matrice » sont précédées d'une pause beaucoup plus longue que les autres subordonnées (Figure 46).

	ddl	F	Sig.
Constante	1	1115,44	p<0,001
numéro de la clause dans la phrase	1	67,54	p<0,001
matrice / non matrice	1	4,48	p=0,035
obligatoire / facultative	3	5,18	p=0,001
type syntaxique _ 2	5	4,40	p=0,001

Tableau 38. Pause initiale de la clause - Effets fixes du modèle 2

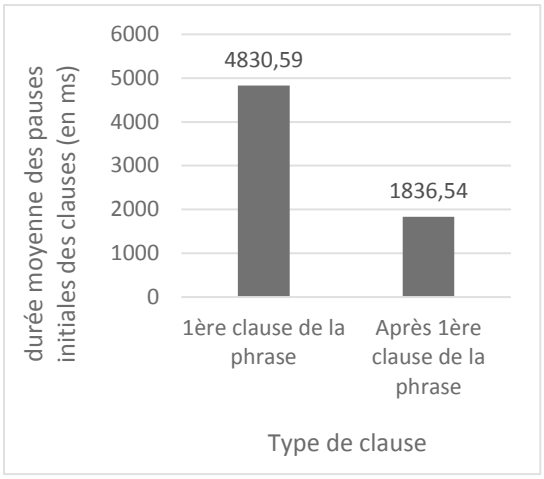


Figure 43. Durée de la pause initiale de la clause en fonction du numéro de la clause dans la phrase (moyennes en ms)

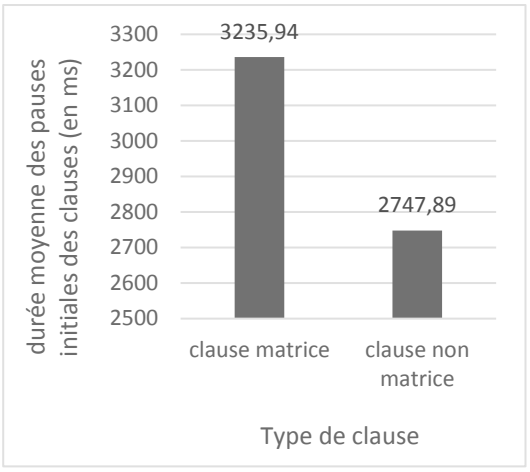


Figure 44. Durée de la pause initiale de la clause en fonction de son aspect matrice ou non matrice (moyennes en ms)

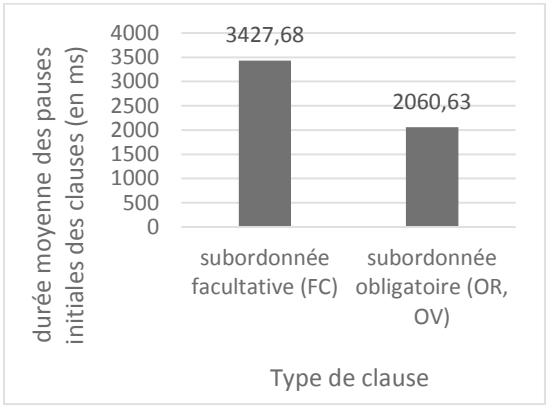


Figure 45. Durée de la pause initiale de la clause en fonction de son aspect obligatoire ou facultatif (moyennes en ms)

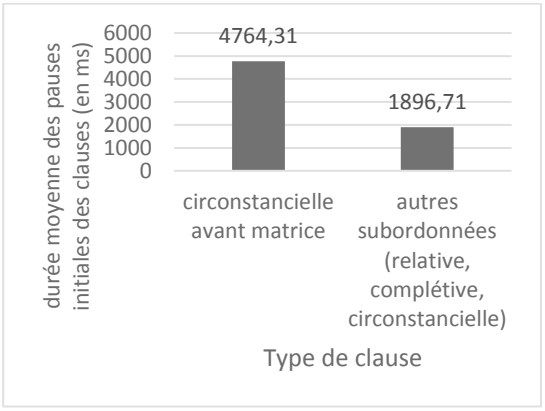


Figure 46. Durée de la pause initiale de la clause en fonction du type syntaxique (moyennes en ms)

2.2.5. Temps d'écriture de la clause en fonction du lien syntaxique entre les clauses

Pour les analyses suivantes, nous avons procédé de la même manière que pour l'analyse des pauses initiales de la clause. Le but est de voir dans quelle mesure le lien de dépendance entre les clauses permet d'expliquer la durée d'écriture de la clause. En effet, une clause déjà planifiée sera sans doute écrite plus rapidement, alors que si pendant l'écriture de la clause la suite est en train d'être planifiée, l'écriture sera plus lente. Nous avons testé des modèles ayant pour effets fixes les facteurs : « numéro de la clause dans la phrase _ 2 », « numéro de la clause de l'unité A _ 2 », « matrice », « obligatoire », « type syntaxique », « profondeur syntaxique », « non finie ». Nous avons testé plusieurs possibilités afin de trouver le modèle apportant l'amélioration maximale par rapport au modèle vide, tout en ayant le moins de paramètres. Pour cela, nous avons parfois regroupé des modalités d'un même facteur, suivant les résultats des tests post-hoc. Pour plus de clarté, nous ne présenterons que l'ICC du modèle vide et le modèle final obtenu, que nous nommerons « modèle 1 ». Des graphiques permettront d'illustrer les différences de moyenne significatives entre les modalités de chaque facteur.

2.2.5.1. Écriture interne de la clause (en ms / seq)

L'ICC du modèle vide est de 16,35%. Le meilleur modèle est celui où sont entrés simultanément les effets principaux des facteurs « numéro de l'unité A dans la phrase » et « profondeur syntaxique », qui distingue les clauses subordonnées des clauses indépendantes. Le modèle 1 est ainsi significativement meilleur que le modèle vide ($\chi_{(7)}$=39,13, p<0,001), les effets fixes ajoutés apportant un gain explicatif de 3,38% .

Ainsi, les clauses indépendantes sont globalement écrites plus lentement que les clauses subordonnées (Figure 47). Les clauses appartenant à la 2ème unité A de la phrase sont écrites plus rapidement que celles appartenant à la 1ère ou à la 4ème unité A de la phrase (Figure 48). Ce modèle permet de valider les hypothèses 2g et 2h.

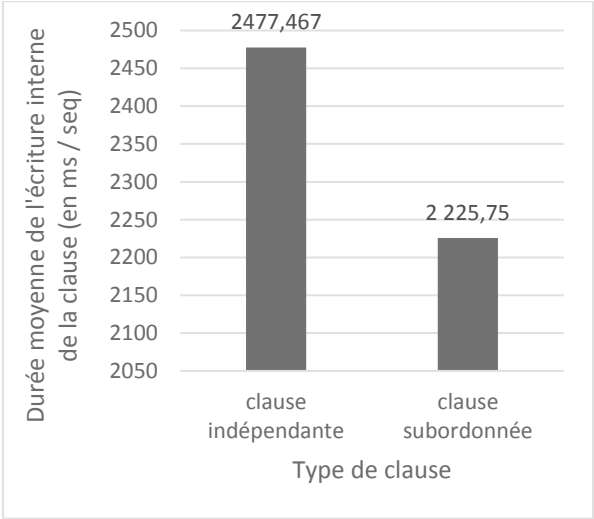


Figure 47. Durée moyenne de l'écriture interne de la clause en fonction de sa profondeur syntaxique (en ms/seq)

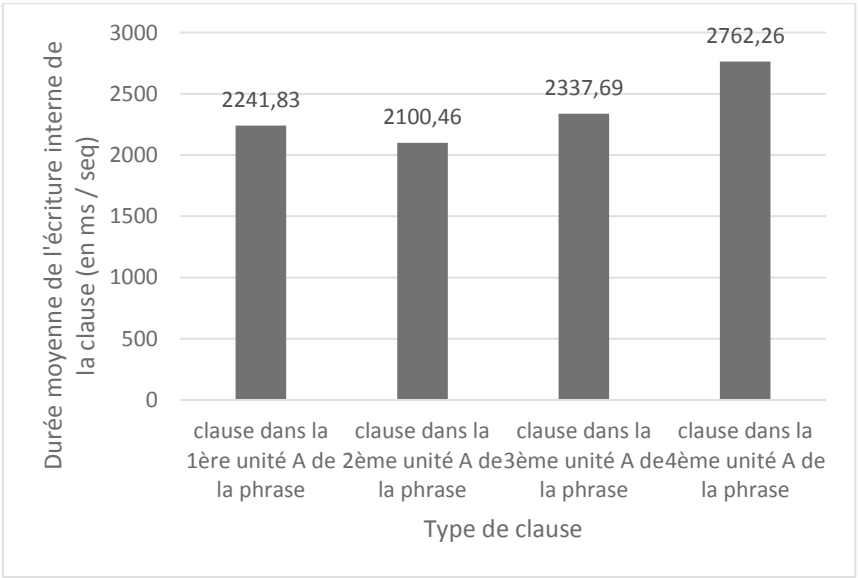


Figure 48. Durée moyenne de l'écriture interne de la clause en fonction du numéro de l'unité A dans la phrase (en ms/seq)

2.2.5.2. Moyenne des pauses inter-mots de la clause (en ms / seq)

L'ICC du modèle vide est de 15,18 %. Le meilleur modèle est celui où sont entrés simultanément les effets principaux des facteurs « numéro de l'unité A dans la phrase _ 2 groupes » et « profondeur syntaxique ». Le modèle 1 est ainsi significativement meilleur que le modèle vide ($\chi_{(5)}$=39,40, p<0,001), les effets fixes ajoutés apportant un gain explicatif de 3,35%.

D'après ce modèle, les clauses indépendantes ont globalement des pauses entre les mots plus longues que les clauses subordonnées (Figure 49). Les « deux groupes » liés au numéro de l'unité A dans la phrase distinguent ici d'une part les clauses appartenant à la 1ère ou à la 2ème unité A de la phrase, et d'autre part celles appartenant à la 3ème ou à la 4ème unité A de la phrase. Les clauses appartenant aux premières unités A ont globalement des pauses entre les mots plus courtes (Figure 50). Ce modèle permet de valider les hypothèses 2g et 2h.

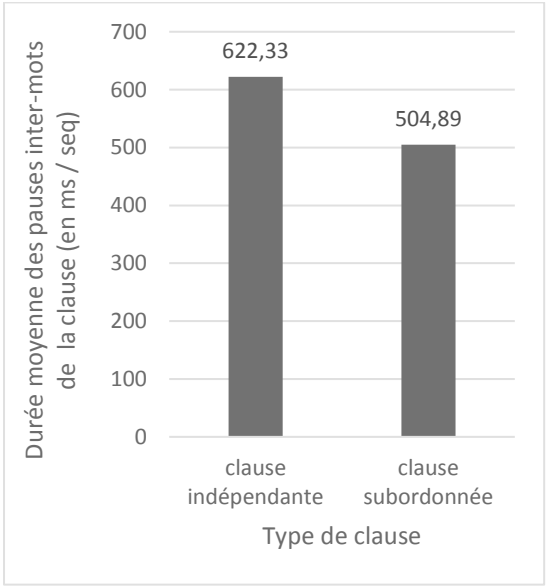


Figure 49. Durée moyenne des pauses inter-mots de la clause en fonction de sa profondeur syntaxique (en ms/seq)

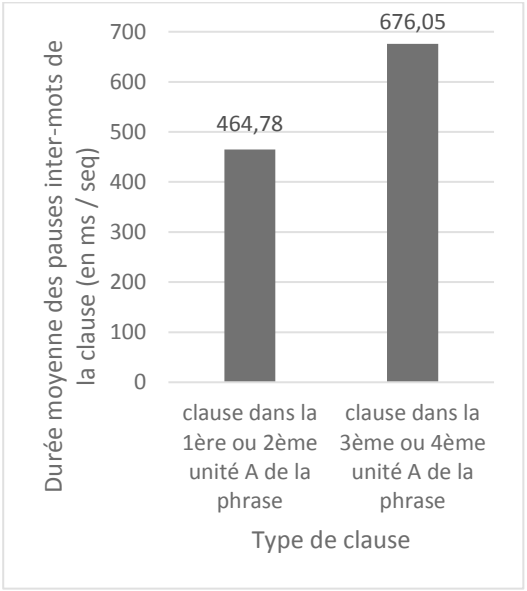


Figure 50. Durée moyenne des pauses inter-mots de la clause en fonction du numéro de l'unité A dans la phrase (en ms/seq)

2.2.5.3. Moyenne des pauses intra-mots de la clause (en ms / car)

L'ICC du modèle vide est de 29,41 %. Le meilleur modèle est celui où sont entrés simultanément les effets principaux des facteurs « numéro de la clause dans la phrase _ 2 » et « profondeur syntaxique ». Le modèle 1 est ainsi significativement meilleur que le modèle vide ($\chi_{(5)}$=35,84, p<0,001), les effets fixes apportant un gain explicatif de 2,23%.

D'après ce modèle, les clauses indépendantes ont globalement des pauses intra-mots plus longues que les clauses subordonnées (Figure 51). Par ailleurs, les pauses intra-mots de la 1ère clause de la phrase sont significativement plus longues que les pauses intra-mots des clauses suivantes (Figure 52). Ce modèle permet de valider les hypothèses 2g et 2h.

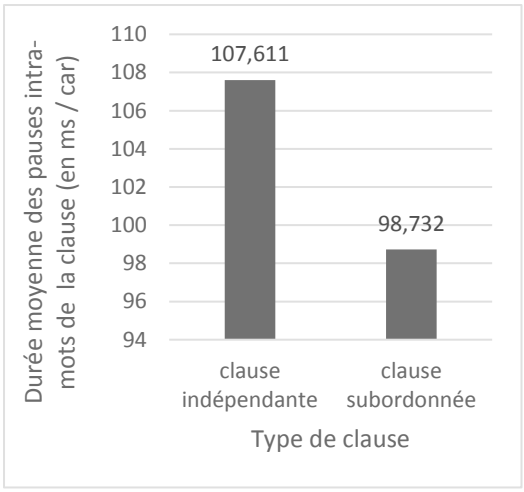


Figure 51. Durée moyenne des pauses intra-mots de la clause en fonction de sa profondeur syntaxique (en ms/car)

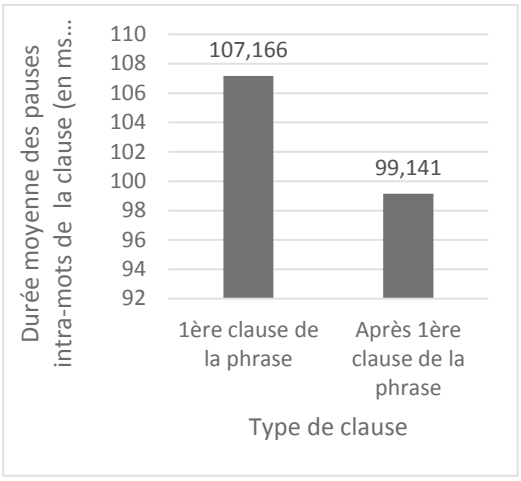


Figure 52. Durée moyenne des pauses intra-mots de la clause en fonction du numéro de l'unité A dans la phrase (en ms/car)

2.2.5.4. Moyenne du temps d'exécution (en ms / car)

L'ICC du modèle vide est de 56,37%, ce qui signifie que plus de la moitié de la variance du temps d'exécution graphique est due à des différences interindividuelles. Pour cette variable dépendante, aucun facteur ne permet de fournir un modèle significativement meilleur que le modèle vide.

En revanche, nous avons regardé ensuite uniquement les clauses subordonnées. Pour celles-ci, le fait de distinguer entre « subordonnées finies » et « subordonnées non finies » apporte un gain explicatif de 2,15%. Le modèle obtenu avec l'ajout de ce facteur est significativement meilleur que le modèle vide ($\chi_{(1)}$=8,18, p<0,005). Ainsi, le temps d'exécution graphique est plus long pour les clauses non finies que pour les clauses finies (Figure 53). Cela va à l'encontre de l'hypothèse 2i.

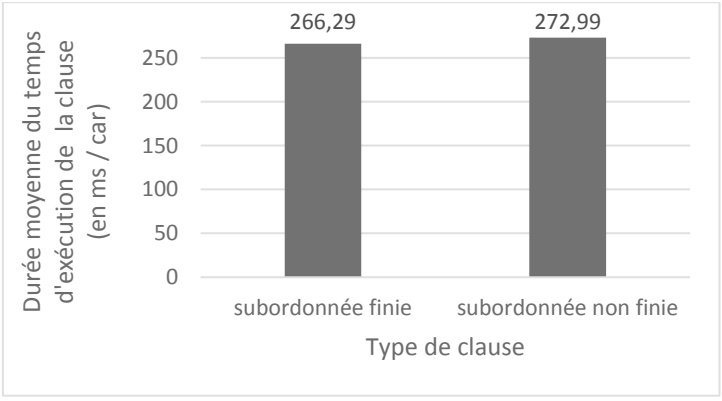


Figure 53. Durée moyenne du temps d'exécution de la clause en fonction du type de subordonnée (en ms/car)

2.2.6. Discussion sur les liens syntaxiques entre les clauses

Le Tableau 39 résume les résultats obtenus grâce aux modèles linéaires mixtes. Toutes les hypothèses avancées n'ont pas été vérifiées par nos analyses. Il faut cependant rappeler que nous avons utilisé des modèles mixtes : par conséquent, il est possible que certains effets ne soient pas significatifs lorsque toutes les variables sont prises en compte simultanément.

Le premier regard porté sur le tableau récapitulatif indique d'une part que les pauses initiales de clauses sont affectées par de nombreuses caractéristiques de ces clauses (numéro de la clause dans la phrase, matrice ou non matrice, nécessité, type syntaxique), et d'autre part que les données chronométriques d'écriture de la clause sont globalement affectées par les mêmes éléments (numéro de l'unité A ou de la clause dans la phrase, profondeur syntaxique).

Les résultats obtenus suggèrent fortement que les clauses subordonnées sont planifiées avant leur pause initiale. En effet, il a été remarqué que les pauses initiales des clauses matrices sont plus longues que celles des non matrices : la pause initiale de la clause matrice est peut-être un des lieux de cette planification anticipée. En outre, l'écriture des subordonnées, que ce soit l'écriture globale, les pauses inter-mots ou les pauses intra-mots, est plus rapide que celle des clauses indépendantes, suggérant qu'une partie de la planification a déjà eu lieu. Le fait que les subordonnées soient anticipées n'est donc pas révélé par nos données au niveau des pauses initiales, comme dans l'étude de Immonen et Mäkisalo (2010), mais par le temps d'écriture de la clause. Cette planification peut se faire à divers degrés : ainsi, les clauses obligatoires d'un point de vue valenciel ou référentiel semblent être davantage planifiées que les clauses facultatives, puisque leur pause initiale est plus courte.

Cependant, les résultats obtenus incitent à être prudent sur la manière dont la planification des clauses est anticipée et son impact sur les données chronométriques. Par exemple, la pause initiale de la phrase ne dépend pas du nombre de clauses ou d'unités A dans la phrase. Cela peut soit s'expliquer par le fait que l'ensemble de la phrase n'est pas planifié en amont, soit parce que la pause initiale de la phrase n'est pas le lieu privilégié de cette planification, et que celle-ci se déroule par exemple pendant l'écriture de la fin de la phrase précédente. Cette dernière hypothèse pourrait peut-être expliquer que les dernières unités A soient écrites plus lentement. Cependant, cette écriture plus lente peut également s'expliquer par une deuxième période de planification au cours de la phrase, si tout n'a pas été planifié au début, ou par une relecture en parallèle de ce qui a déjà été écrit. Les pauses intra-mots étant plus longues pendant l'écriture de la 1ère clause de la phrase, il est possible qu'une partie de la planification ou de la formulation de l'ensemble de la phrase se déroule durant le début de la production de la phrase.

Il convient également de s'attarder sur la variable « temps d'exécution ». Celle-ci est affectée par peu de variables : seules les subordonnées « non-finies » se révèlent être écrites plus lentement que les subordonnées « finies ». Or, nous pourrions penser que les clauses non-finies sont davantage anticipées puisque le verbe n'est pas conjugué avec un verbe propre à cette clause. L'explication pourrait alors être que ces subordonnées sont souvent considérées comme étant plus difficiles à traiter : elles sont très peu utilisées par les enfants, alors que ce sont des structures fréquemment employées par les adultes (Jisa, 2000) ; de plus, chez ces derniers, ce sont des structures davantage utilisées à l'écrit (Gayraud et al., 2001). La caractéristique « non finie » semble donc marquer une difficulté plutôt qu'un lien de dépendance plus étroit.

Variable dépendante	Clauses analysées	Facteurs permettant d'expliquer au mieux la variance	Résultats
Pause initiale (en ms)	Toutes les clauses	Numéro de la clause dans la phrase _ 2	1ère clause > clauses suivantes
		Matrice / non matrice	Matrice > non matrice
		Obligatoire / facultative	FC > OR et OV
		Type syntaxique	Circonstancielle avant matrice> autres subordonnées (relative, complétive, circonstancielle)
Écriture interne (en ms / seq)	Toutes les clauses	Numéro de l'unité A dans la phrase	n°2 < n°1 et n°2 < n° 4
		Profondeur syntaxique	Indépendante> subordonnée
Pauses inter-mots (en ms / seq)	Toutes les clauses	Numéro de l'unité A dans la phrase	n°1 et 2 < n° 3 et 4
		Profondeur syntaxique	Indépendante > subordonnée
Pauses intra-mots (en ms / car)	Toutes les clauses	Numéro de la clause dans la phrase	n°1 > clauses suivantes
		Profondeur syntaxique	Indépendante > subordonnée
Temps d'exécution (en ms / car)	Toutes les clauses	(rien de significatif)	
	Clauses subordonnées	Finies / non finies	Non finies > finies

Tableau 39. Facteurs liés au lien entre les clauses expliquant le mieux la variance des données chronométriques des clauses

3. Conclusion

Ce chapitre nous a permis d'explorer de manière préliminaire les données chronométriques du corpus "Dimitri". Des résultats présents dans les études précédentes ont été corroborés, tels que l'effet de l'emplacement syntaxique en fonction de la durée de la pause initiale ou l'impact du degré de dépendance sur les données chronométriques. Il apparait alors que la segmentation en phrases, unités A et clauses soit pertinente, puisque des processus différents semblent être engagés à ces divers niveaux. Cependant, même si les résultats suggèrent que la planification des clauses soit anticipée, ils restent très difficiles à interpréter. En effet, nous ne sommes pas certains que des pauses initiales ou des temps d'écriture longs soient systématiquement synonymes de préparation (conceptuelle ou linguistique) de ce qui suit. Pour explorer cette hypothèse, il reste donc à identifier à quels moments les différentes composantes du récit sont préparées. Cette question, qui est au centre de notre travail, sera traitée dans le chapitre qui suit.

Les résultats obtenus jusqu'à présent nous permettent d'affiner la manière dont les analyses seront conduites. Tout d'abord, de nombreuses différences interindividuelles ont été relevées, notamment quant au temps moyen d'exécution et à la longueur des pauses initiales. Il conviendra donc d'utiliser un modèle prenant en compte la variable « participants ». Par ailleurs, la phrase apparait comme une possible unité de production, bien que la longueur très élevée (presque 10 s) de la pause initiale nous interroge sur les processus cognitifs engagés à cet emplacement. Il faudra donc mener des analyses spécifiquement sur le niveau syntaxique « phrase » afin d'identifier dans quelle mesure cette unité est un tout planifié dès le début de sa production. Enfin, l'étude de l'impact des liens syntaxiques entre les clauses sur les données chronométriques a fait ressortir des différences entre clauses subordonnées et non subordonnées, et matrices et non matrices. Pour plus de précisions, il sera donc pertinent d'étudier la préparation des clauses en distinguant ces différents types de clauses.

Chapitre 7 - Lien entre rythme d'écriture et contenu linguistique des unités

Afin de rédiger la suite d'un récit, le scripteur doit effectuer de nombreux choix. Au niveau macro-structural, il doit affiner son scénario (e.g. principaux évènement, dénouement final, personnages introduits) ; au niveau micro-structural, il doit choisir quels éléments précis introduire (prédicat, arguments du prédicat). Enfin, il faut choisir une structure syntaxique et les lexèmes adéquats. Ces choix sont en partie contraints par les règles de la langue, et par le texte déjà-écrit : il faut par exemple que le texte soit cohérent, que les choix lexicaux respectent la continuité référentielle. Certaines décisions peuvent être prises lors de la pause de pré-écriture, mais il est vraisemblable que de nombreux choix soient faits au fur et à mesure de l'écriture du texte. Le chapitre précédent a suggéré que la phrase était en partie planifiée dès le début de son écriture et que les clauses subordonnées étaient vraisemblablement planifiées dès l'écriture de la clause matrice. Deux questions se posent alors afin de préciser cette gestion temporelle :

- À quel moment les divers éléments linguistiques sont-ils planifiés[60] ?
- Certaines caractéristiques (fréquence, longueur, accessibilité) rendent-elle la planification plus couteuse, ou au contraire plus facile ?

Pour répondre à ces questions, de nombreux éléments sémantiques, syntaxiques et lexicaux ont été codés dans notre corpus. Dans un premier temps, nous présenterons le codage utilisé pour décrire le contenu linguistique des phrases et des clauses ; nous présenterons ensuite le déroulement général des analyses. Dans un second temps nous lierons les caractéristiques linguistiques relevées avec les mesures chronométriques rattachées aux phrases et aux clauses. Ainsi, il nous sera possible de comprendre de manière plus fine la gestion temporelle locale des processus de micro-planification et formulation.

[60] Tout comme dans les chapitres précédents, le terme « planifié » est utilisé par commodité pour désigner la réflexion portant sur les choix sémantiques locaux, syntaxiques et lexicaux ; ils regroupent donc la micro-planification et la formulation)..

1. Méthodologie

1.1. Codage du contenu linguistique

Bien qu'il soit impossible de faire une description exhaustive de chaque unité, nous avons souhaité décrire chaque clause de la manière la plus précise possible, tant sur le plan sémantique (e.g. quel type d'action est effectué, quels compléments circonstanciels sont mentionnés,) que syntaxique (e.g. quelles formes verbales sont utilisées, quelle est la complexité syntaxique de la clause) ou lexicale (e.g. quelle est la fréquence moyenne des lexèmes utilisés, quelle est la nature grammaticale choisie pour exprimer le sujet). Il est nécessaire, tout d'abord, d'expliciter la segmentation utilisée au sein de la clause et la manière dont nous avons codé les syntagmes. Nous présentons ensuite le codage sémantique des unités, dont un certain nombre de choix sont spécifiques à notre corpus. Pour établir certaines définitions ou faire des choix terminologiques, nous nous réfèrerons souvent au Dicovalence, qui est un dictionnaire électronique développé par van den Eynde et Mertens (2010), répertoriant les informations liées à la valence de plusieurs milliers de verbes français.

1.1.1. Segmentation sous-clausale : syntagmes, mots, lexèmes

Afin d'être plus précis lors de la description linguistique des phrases, unités A et clauses, les textes ont été segmentées en unités inférieures à la clause : syntagmes, mots et lexèmes.

1.1.1.1. Les syntagmes

Lors de ce travail, les textes ont été segmentés en constituants immédiats, que nous nommerons « syntagmes ». Afin de décrire les relations entre les syntagmes, il est nécessaire de considérer qu'il y a plusieurs « niveaux », c'est-à-dire que des syntagmes vont être emboités dans d'autres. Parfois, la structure de surface ne fait pas apparaitre ces emboitements, mais parfois ils sont clairement visibles, comme lorsque nous parlons de « subordonnées enchâssées », comme dans l'exemple (59).

(59) il jeta le livre ***[qu'il avait soigneusement dissimulé]*** à *l'autre bout de sa cellule* [S02]

Les syntagmes sont en relation avec le prédicat, qui sert de pivot à la clause. Le niveau supérieur est l'unité A ; les constituants s'articulent ensuite autour du prédicat de la clause matrice de l'unité A. Chaque syntagme peut ensuite lui-même compter plusieurs syntagmes. Nous nous sommes cependant limités à segmenter les unités dans la mesure où elles gardaient une autonomie sémantique ou référentielle. Par exemple, au sein d'un syntagme nominal, nous n'avons pas séparé le déterminant du nom ; au sein du groupe verbal, la négation n'a pas été isolée ; les expressions figées n'ont pas été segmentées ; les prépositions et conjonctions de subordination n'ont pas été isolées.

La phrase (60) ci-dessous illustre cette segmentation, chaque partie entre crochets étant un syntagme. Les indices indiquent le niveau de profondeur syntaxique ; les crochets dont l'indice est « 0 » délimitent l'unité A.

(60) **[[Il]$_1$ [**ouvrit**]$_1$ [**la porte de sécurité **[[**que**]$_3$ [**le gardien**]$_3$ [***n'avait pas refermée***]$_3$]$_2$]$_1$]$_0$** [S16]

Niveau 1	Niveau 2	Niveau 3
il		
ouvrit		
la porte de sécurité que le gardien n'avait pas refermée	que le gardien n'avait pas refermée	que
		le gardien
		n'avait pas refermée

Tableau 40. Codage de la profondeur syntaxique des syntagmes - Exemple

Comme le montre le Tableau 40, l'unité A comporte trois syntagmes au niveau 1 : le prédicat « ouvrit », le sujet « il » et le complément d'objet direct « la porte de sécurité que le gardien n'avait pas refermée ». Le complément d'objet est un syntagme nominal étendu d'un syntagme qui est une clause (de niveau 2) ; cette clause est elle-même composée de trois syntagmes (de niveau 3).

Une information sur la profondeur syntaxique de chaque syntagme a donc été donnée : par exemple, la clause « que le gardien n'avait pas refermée » est de niveau 2. Cette information sur la profondeur syntaxique permet d'être plus précis lors de la description du lien entre les clauses subordonnées et la clause matrice d'une unité A.

Soit l'exemple (61), pour lequel on indique uniquement la segmentation en clauses :

(61) une fois la voie dégagée **/** il sortit par la porte **/** qui donnait *sur l'arrière*-cour [S28]

Dans cette unité A, il y a une clause matrice dont le prédicat est « sortit », puis deux clauses qui lui sont subordonnées : une circonstancielle dont le prédicat est « dégagée », et une relative dont le prédicat est « donnait ». On pourrait ainsi dire qu'il y a deux « niveaux » de clauses (le niveau de la matrice et le niveau des deux subordonnées).

L'exemple (62) reprend la même unité A, mais avec la segmentation en syntagmes.

(62) [[une fois [la voie]$_2$ [dégagée]$_2$]$_1$ [il]$_1$ [sortit]$_1$ [par la porte [[qui]$_3$ [donnait]$_3$ [sur *l'arrière*-cour]$_3$]$_2$]$_1$]$_0$ [S28]

Ici, on voit que le prédicat de la clause circonstancielle « une fois la voie dégagée » est de niveau 2, alors que celui de la clause relative « qui donnait sur l'arrière-cour » est de niveau 3.

1.1.1.2. Les mots

Nous utilisons ici le terme de « mot » pour désigner une unité sémantique. Par conséquent, ont été comptabilisés comme un seul mot :

- les mots composés : e.g. « porte d'entrée », « plateau-repas ».

Dans ces cas, le sens n'est pas obtenu par l'addition des sens de chaque mot. D'un point de vue sémantique, seul un référent est mis en jeu (par exemple, il n'est pas possible d'insérer un élément entre les termes du mot composé).

- les locutions
  - adverbiales : e.g. « à peine »,
  - prépositionnelles : e.g. « à travers »
  - conjonctives : e.g. « dès que »
- les deux parties de la négation : « ne » et « pas »
- les verbes pronominaux : e.g. « se coucher », s'élancer »
- les verbes à un temps composé : e.g. « avait alerté »

Ainsi, la phrase (63) comporte 13 mots, séparés par « _ ».

(63) *Son _ cœur _ faillit _ s'arrêter _ lorsque _ le _ gardien* _ qui _ gardait _ la _ *porte d'entrée* _ *l'* _ *interpella.* [S29]

Par ailleurs, chaque mot a été codé comme étant soit un mot grammatical, soit un mot lexical, selon la répartition indiquée dans le Tableau 41.

Mots grammaticaux	Mots lexicaux
Conjonctions	Nom communs
Pronoms	Nom propres
Déterminants	Verbes
Prépositions	Adjectifs
Certains adverbes[61]	Interjections
	Certains adverbes

Tableau 41. Mots grammaticaux et lexicaux

Dans le cas des locutions, la catégorie du « mot » correspondait à la catégorie de la locution (ex : « tout à fait » est un mot grammatical, et non deux mots grammaticaux et un mot lexical). Pour les verbes, ils ont été comptabilisés comme un seul mot lexical (ex : « se coucher » est un mot lexical, et non un mot grammatical et un mot lexical).

[61] Les adverbes se finissant en « -ment », et dont le sens est déductible de la dérivation, ont été comptabilisés parmi les mots lexicaux ; les autres adverbes sont considérés comme des mots grammaticaux

1.1.1.3. Les lexèmes

Les lexèmes sont les « séquences » (chaines de caractères séparées par des espaces blancs) équivalant à un mot présent dans le dictionnaire, soit ce que l'on pourrait appeler un « mot graphique ». Par exemple, la locution adverbiale « à peine » compte deux lexèmes.

1.1.2. Codage des syntagmes

Pour chacun des syntagmes, il est possible d'indiquer sa nature, ainsi que d'apporter des informations d'ordres syntaxique et sémantique. Le but de ce travail n'est pas de faire une analyse linguistique exhaustive de ces syntagmes, mais de voir dans quelle mesure certaines de ces caractéristiques peuvent influencer le processus de production. Quatre indications sont données pour chaque syntagme : sa nature, son rôle syntaxique, son rôle sémantique, l'aspect obligatoire ou facultatif de sa présence.

1.1.2.1. Nature du syntagme

Les syntagmes peuvent être de différentes natures. Le Tableau 42 illustre les natures grammaticales relevées dans le texte, et indique les abréviations utilisées. La nature dont il est question dans la ligne est mise en gras dans l'exemple.

Nature grammaticale	Spécification de la nature grammaticale	Exemple
Clause		(64) Elle coupa un passage dans le grillage **pour laisser sortir son mari** [S04]
Verbe	Verbe simple (« V »)	(65) Elle **coupa** un passage dans le grillage pour laisser sortir son mari [S04]
	Locution verbale (« LV »)	(66) Dimitri **prit ses jambes à son cou** [S12]
Nom	Nom propre (« NP »)	(67) **Dimitri** prit ses jambes à son cou [S12]
	Syntagme nominal simple (« SN »), avec ou sans préposition	(68) Elle coupa un passage **dans le grillage** pour laisser sortir **son mari** [S04]
	Syntagme nominal étendu (« SNE »)	(69) Et enfila **les vêtements du gardien** [S13] (70) Il sortit ***par la porte qui donnait sur l'arrière-cour***. [S28]
Pronom (« PRO »)		(71) **Il** avait pris la matraque **que** sa fiancée avait réussi à **lui** faire passer [S22] (72) *Je n'ai pas le temps de m'occuper* **de toi**. [S03]

Nature grammaticale	Spécification de la nature grammaticale	Exemple
Adjectif	Simple (« ADJ »)	(73) Il repensa encore à ce complice **inconnu** [S12]
	Syntagme adjectival étendu (« ADJE »)	(74) plongea la tête du gardien dans le mouchoir **imbibé de chloroforme** [S07]
Adjectif possessif (« POS »)		(75) Il sortit de **sa** cellule [S02]
Adverbe (« ADV »)		(76) Il sortit **ensuite** de sa cellule [S01] (77) Il expira **fortement** [S22]
Conjonction (« CONJ»)		(78) **Et** *s'en alla [S21]*[62]
Interjection (« ITJ »)		(79) **Stop** ! [S12]

Tableau 42. Natures grammaticales relevées pour les syntagmes du corpus

Nous regrouperons sous le terme « syntagmes nominaux » les SN, SNE, PRO et NP.

1.1.2.2. Rôle syntaxique

Les syntagmes peuvent avoir différentes fonctions au sein de la clause. Nous allons présenter les informations syntaxiques que nous avons codées, en indiquant à chaque fois par quels types de syntagmes elles peuvent être accomplies. Notre codage concernant les rôles syntaxiques a été effectué autour de la notion de « prédicat », utilisé pour délimiter nos clauses. Il s'agira donc de voir quel rôle est joué par chaque syntagme par rapport à ce prédicat.

Prédicat « PRED »

Le prédicat est le « noyau » de la clause. Nous pourrions souvent associer celui-ci à un syntagme « verbal », mais nous allons voir que parfois ce syntagme prédicatif dépasse le simple verbe. La notion de prédicat permet ici de caractériser les syntagmes qui sont les pivots de la clause ; le prédicat peut donc être de différentes natures.

`Les verbes conjugués`

Pour définir ce que nous avons considéré comme un prédicat verbal, nous pouvons nous reporter au chapitre « Méthodologie générale », dans lequel nous avons explicité notre découpage en clauses, et donc les cas où des verbes étaient considérés comme un unique prédicat, ou au contraire des prédicats séparés. Ces définitions permettent d'identifier les prédicats correspondant à des verbes.

Cependant, il existe des cas où la clause semble être centrée sur un verbe, alors que le réel prédicat n'est pas ce verbe, comme cela va être présenté dans les sections suivantes.

[62] Nous avons suivi la proposition de (Riegel et al., 2002) de classer « donc » parmi les adverbes de liaison, et non parmi les conjonctions de coordination, en raison de son comportement syntaxique.

Les locutions verbales

Il existe de nombreux cas où nous avons l'impression que le prédicat est un verbe, alors qu'en réalité le noyau de la clause est une expression figée, ou un nom accompagné d'un verbe support. Dans ces cas, il s'agit donc plutôt de locutions verbales. Ce constat n'influence pas le découpage en clauses, mais intervient dans la délimitation du syntagme considéré comme « prédicat ».

En suivant la définition de Gross (1993) qui s'intéressait aux phrases figées, on peut définir une locution verbale comme étant un assemblage « verbe + complément », dont le sens n'est pas compositionnel ; il est notamment difficile de pronominaliser le complément. Les verbes supports sont très souvent « avoir », « prendre » et « faire ». Des dictionnaires, tels que le Trésor de la Langue Française, répertorient de telles locutions verbales.

L'Annexe 2 présente les séquences de notre corpus que nous avons considérées comme locutions verbales, comme par exemple « avoir confiance » ou « prendre garde ».

Les adjectifs, adverbes ou noms accompagnés d'une copule

Il arrive fréquemment que le prédicat de la clause ne soit pas un verbe, mais par exemple un adjectif ou un adverbe : c'est le cas lorsque le verbe est une simple copule. En français, une phrase grammaticale doit de manière prototypique contenir un verbe ; par conséquent, si le prédicat est un adjectif, il est nécessaire de rajouter une copule (le verbe « être »). Dans la majorité des cas, le prédicat est alors un adjectif (80), mais il arrive que ce soit un adverbe (81) ou un syntagme nominal (82).

(80) Il **était libre. [S11]**

(81) Irina **était là. [S2]**

(82) Tout cela ***n'était qu'un magnifique rêve*** **[S10]**

Prédicats incluant un pseudo-paradigme

Lors de l'analyse en syntagmes de la phrase, les pronoms sont généralement isolés car ils remplissent une certaine fonction syntaxique, par exemple sujet ou oblique. Cependant, il arrive que des pronoms ne remplissent pas véritablement cette fonction : dans le Dicovalence, ils ont été identifiés comme « pseudo-paradigmes ». Dans ces cas, nous ne les avons pas isolés, ils font donc partie du prédicat. Les cas les plus fréquents sont le pseudo-il (83) et le pseudo-se (84) ; on peut aussi trouver des pseudo-y, pseudo-en, etc.

(83) **Il fallait** *qu'il quitte cet endroit [S29]*

(84) Par surprise il **se jeta** sur le geôlier [S28]

Sujet du prédicat « SUJ »

Il s'agit du syntagme avec lequel le verbe conjugué du prédicat s'accorde en genre et en nombre (85).

(85) **Il** traversa la cour [S11]

Il arrive que le sujet ne soit pas réalisé par un élément lexical, dans le cas d'ellipses (86) ou de clauses non finies (87).

(86) Il parcourut à peine six mètres / et **[ ]** *trouva l'emplacement idéal*. // [S28]

(87) il demanda alors à son ami le gardien / de **[ ]** faire diversion [S05]

Le sujet est le plus souvent un syntagme nominal : il peut s'agir d'un nom, d'un SN, d'un SNE ou d'un pronom. En théorie une clause peut également avoir la fonction de sujet (88), mais un tel cas n'est pas présent dans notre corpus.

(88) **Courir le plus vite possible** *était indispensable pour pouvoir s'évader.*

Compléments du prédicat

`Compléments de verbe`

Ces compléments permettent de saturer la valence du verbe. Nous relevons les compléments d'objet direct, complément d'objet second et complément adpositionnels.

- Compléments d'objet (direct) du verbe « OBJ» (89) :

(89) Il traversa **la cour** [S11]

Il peut s'agir d'un syntagme nominal (NOM, SN(E), PRO) ou d'une clause.

- Compléments d'objet second « OBJ2 » (90) : Nous avons désigné ainsi les clauses non-finies pour lesquelles il y a une relation de montée ou de contrôle avec l'objet direct du verbe de la clause matrice.

(90) *alors qu'il entendait les pas du gardien* **se rapprocher de son lit** [S25]

- Complément adpositionnel « ADP » (91) : Il s'agit de complément introduit par une préposition, et qui font partie de la valence du verbe.

(91) Dimitri **lui** ouvrit la porte de sa cellule [S15]

(92) Il sortit **de sa cellule** [S02]

Ces compléments sont réalisés par des syntagmes nominaux introduits par une préposition ou par des clauses.

Complément de nom « CNP »

Ces syntagmes complètent un nom qui fait partie d'une locution verbale, ou qui est le prédicat de la clause.

(93) Il était désormais temps ***qu'il quitte à son tour ses vêtements*** [S11]

Il peut s'agir d'une clause (93), mais également d'un syntagme nominal, d'un syntagme adjectival ou d'un adverbe.

Compléments d'adjectifs « CAP »

Ces syntagmes complètent un adjectif qui est lui-même prédicat de la clause ; il s'agit généralement de clauses (94).

(94) Il était beaucoup plus prudent **de continuer à pieds** [S04]

Compléments circonstanciels « CIR »

Ces compléments donnent des informations sur les circonstances dans lesquelles le prédicat est réalisé (temps, lieu, manière, etc.) Ils peuvent être réalisés par des syntagmes nominaux (95), des adverbes (96) ou des clauses (clauses circonstancielles (97)).

(95) *Il n'avait rencontré aucun garde* **sur sa route** [S14]

(96) **Là**, il se regarda dans le miroir [S13]

(97) *et s'enfuit* **en courant** [S24]

Compléments internes

Certains syntagmes se situent à l'intérieur d'un des syntagmes mentionnés ci-dessus (prédicat, sujet du prédicat, complément du prédicat ou complément circonstanciel). Ils ne jouent pas directement un rôle par rapport au prédicat, mais servent à compléter un élément qui n'est pas un prédicat : c'est pourquoi nous les classons sous la rubrique « complément interne ».

Complément de nom « CN »

Les compléments de nom (« CN ») peuvent être réalisés par des clauses (complétives (98), relatives (99)), des syntagmes adjectivaux (100) ou des syntagmes nominaux (101). Nous avons également pris en compte les possessifs, car d'un point de vue sémantique il est possible de rapprocher les exemples (101) et (102).

(98) *qui n'avait plus de moyens* **de se défendre** [S02]

(99) Sa main gauche était toujours bandée depuis la dernière altercation ***qu'il avait eue*** **avec Dimitri** [S03]

(100) Dimitri arpente les conduits **étroits** [S08]

(101) saisit les clés **du gardien** [S14]

(102) Dimitri en profita pour lui voler **ses** clés [S07]

`Complément d'adjectif « CA »`

(103) Tous deux, tellement heureux **de se retrouver** *s'enlacèrent en pleurant [S07]*

Les compléments d'adjectifs (« CA ») peuvent être des clauses (complétives non finies) (103) ou des syntagmes nominaux.

Éléments servant de liaison « LIA »

Certains syntagmes jouent un rôle non pas au niveau de la clause, mais au niveau du texte : ils servent à en préciser l'articulation. Il s'agit des conjonctions de coordination et des adverbes de liaison.

(104) **et** *la porte s'entrouvrit en silence* [S21]

1.1.2.3. Informations sémantiques

Il est difficile d'établir une liste exhaustive de tous les rôles sémantiques que peuvent avoir des syntagmes nominaux (par exemple : agent, patient, expérienceur, siège, thème, destinataire, locatif, instrumental, temporel). En outre, une description aussi fine des rôles sémantiques n'est pas nécessairement pertinente dans le cadre de nos analyses. Nous avons simplement donné quelques indications pour certains types de syntagmes.

Compléments de verbe

Les compléments sujet et objet n'ont pas été caractérisés par leur rôle sémantique : ils auraient pu être simplement qualifiés de « Agentif » ou « Patientif », selon le fait qu'ils se comportent de manière similaire à un agent ou un patient prototypique ; cette information est cependant fortement redondante avec le rôle syntaxique, sauf dans les quelques cas de phrases au passif. Nous n'avons donc pas indiqué quel était le rôle sémantique du sujet du prédicat et de l'objet du verbe.

Parmi les autres compléments de verbe, nous avons distingué :

- Le destinataire ou bénéficiaire « DEST » (105) :

(105) Le gardien était venu **lui** apporter son plateau-repas du soir [S17]

Il s'agit des compléments dont la pronominalisation suit le paradigme « lui / leur ».

- Le lieu (106) :

(106) Il sortit **de sa cellule** [S02]

- Les autres obliques « OBQ » (107) :

(107) *Je n'ai pas le temps de m'occuper* **de toi** [S03]

Compléments circonstanciels

Le rôle sémantique des compléments circonstanciels a été renseigné, comme illustré dans le Tableau 43.

Sémantique	Abréviation	Exemple
Temps balisant l'histoire (i.e., introduisant un nouveau temps référentiel)	« BAL »	(108) **Alors** Dimitri comprit [S04]
Temps- autre	« TPS »	(109) Il avait eu beaucoup de chance ***jusqu'à maintenant*** [S05]
Lieu – direction	« DIR »	(110) Et fonça **vers la porte** [S06]
Lieu – localisation	« LOC »	(111) Il fait nuit noire **dans ces conduits** [S08]
Lieu – source	« SOU »	(112) *Bientôt il fut sorti de l'enceinte de la prison [S07]*
Manière	« MAN »	(113) et marcha ***d'un pas ferme*** vers la sortie [S12]
But	« BUT »	(114) Dimitri regarda partout pour essayer de trouver une solution [S26]
Conséquence	« CONS »	(115) Tout avait été si longuement préparé **que le plan ne pouvait faillir** [S22]
Cause	« CAU »	(116) Il pouvait profiter du déménagement de la bibliothèque **puisque tous croyaient *qu'il était déjà* bien loin** [S19]
Instrumental	« INST »	(117) Il le bâillonna **avec un bout de drap** [S02]
Condition	« COND »	(118) ***S'il voulait agir,*** *c'était maintenant ou jamais* [S07]
Comparaison	« COMP »	(119) *Elle s'était imposée à lui* **comme une révélation** [S18]
Comitatif	« COMI »	(120) Viens **avec nous** ! [S10]
Agent	« AG »	(121) Il devait aller chercher sa fiancée enfermée **par le roi** [S17]

Tableau 43. Rôle sémantique des compléments circonstanciels

Mots de liaison

Les connecteurs relevés et le classement effectué par rapport à leur apport sémantique sont indiqués dans le Tableau 44.

Sémantique	Abréviation	Exemple
Addition	« ADD »	et, d'ailleurs
Opposition / concession	« OPP »	mais, tout de même, pourtant, or, cependant
Hypothèse	« HYP »	au cas où
Conséquence	« CONS »	donc
Alternative	« ALTER »	ou
Cause, explication	« CAU »	en effet
Modalisateur	« MOD »	heureusement, surtout, malheureusement, bel et bien, par chance, certainement

Tableau 44. Classement sémantique des mots de liaison

1.1.2.4. Caractère obligatoire des syntagmes

Plusieurs conditions peuvent rendre la présence d'un syntagme « obligatoire », comme cela a été brièvement évoqué dans le chapitre précédent.

D'un point de vue de la valence

Certains syntagmes servent à saturer la valence d'un autre élément. Certes, ils peuvent parfois être omis : par exemple, le complément de lieu « sous le lit » (122) pourrait être omis. Cependant, la valence du verbe « se cacher » indique que ce verbe est généralement accompagné d'un complément de lieu. En revanche, la valence du verbe « se changer » ne comprend pas de complément de lieu : dans l'exemple (123), le complément de lieu est donc facultatif.

(122) *Dimitri s'était caché* **sous le lit**. [S19]

(123) **Dans le camion**, Dimitri se changea donc [S19]

Afin de déterminer si un complément faisait partie de la valence d'un verbe, nous nous sommes systématiquement référés au Dicovalence.

Concernant la valence verbale, le sujet, l'objet et le complément adpositionnel sont considérés comme obligatoires. À l'inverse, les compléments « circonstanciels » sont facultatifs. « Obligatoire » d'un point de vue de la valence ne signifie donc pas que l'absence du complément rendrait la phrase agrammaticale, mais indique que la présence du complément peut être requise par la valence-même du verbe.

Par ailleurs, certains compléments d'adjectifs sont nécessaires pour saturer la valence de l'adjectif (124) ; alors que dans d'autres cas ils pourraient être omis (125).

(124) Mais Amin était prêt **à toutes les luttes** pour sauver son frère Dimitri de la peine de mort [S01]

(125) Abasourdi et fatigué **de sa journée** le gardien ne put résister [S26]

Les compléments d'adjectifs se distinguent donc par leur présence obligatoire ou facultative : si l'omission du syntagme rend la phrase incorrecte d'un point de vue syntaxique, le syntagme est considéré comme « OV » (« obligatoire d'un point de vue valenciel ») ; sinon, il est codé « FV » (« facultatif d'un point de vue valenciel »).

D'un point de vue référentiel

Certains compléments de nom servent à déterminer le nom afin d'identifier le référent, apportant ainsi une restriction quant au choix du référent. Dans ce cas ils sont obligatoires (« OR »).

(126) Il referma la porte **de sa cellule** [S22]

Dans d'autre cas, le CN ajoute un commentaire à propos du référent sans être nécessaire à sa délimitation ; il est alors facultatif (« FR »).

(127) *Il n'y aurait donc qu'une* **vieille** lampe à pétrole [S22]

Cela permet en outre de distinguer deux types de relatives, traditionnellement nommées « restrictives » (« OR ») (56) et « explicatives » (« FR ») (129).

(128) Sa main gauche était toujours bandée depuis la dernière altercation ***qu'il avait* eue avec Dimitri** [S03]

(129) Et ils partirent tous les deux en bateau pour la Crête **où Dimitri possédait une ferme léguée par son parrain** [S19]

D'un point de vue communicatif

Dans la Grammaire méthodique du français (Riegel et al., 2002, p. 144), deux fonctions sont distinguées pour les compléments circonstanciels : la « fonction scénique », servant « à la mise en place préalable du cadre de circonstances où se situe le reste de la phrase » et la « fonction rhématique », lorsque les circonstants « jouent le rôle de propos par rapport au thème que constitue le reste de la phrase ». Dans le premier cas, les syntagmes seraient plutôt facultatifs d'un point de vue communicatif (« FC ») (130), et dans le second cas ils seraient obligatoires (« OC ») (131).

(130) Dimitri se déshabilla **à toute vitesse** [S21]

(131) *Il finit par s'endor*mir et se réveilla **dans sa cellule**. [S10]

Cette distinction peut également s'appliquer aux compléments de nom, qui peuvent parfois être rhématiques (132).

(132) Un **long** *voyage s'annonçait [S01]*

Cela permet également d'isoler un troisième type de relative : les relatives narratives, qui se réfèrent à un évènement qui a un lien temporel avec l'unité phrastique qui précède (57).

(133) Dimitri assomma le gardien **qui tomba évanoui** [S15]

Codage

Un syntagme obligatoire d'un point de vue syntaxique ou référentiel l'est nécessairement d'un point de vue communicatif. Afin de rendre le codage plus concis, un syntagme était noté « OV » ou « OR » s'il était obligatoire d'un point de vue valenciel ou référentiel (suivant s'il complétait un verbe, un adjectif ou un nom) ; si ce n'était pas le cas, il était noté « OC » s'il était obligatoire d'un point de vue communicatif ; si ce n'était pas le cas, il était noté « FC ».

1.1.2.5. Récapitulatif

Le Tableau 45 permet de résumer notre codage concernant les syntagmes de la phrase. Nous y retrouvons les rôles syntaxiques pouvant être joués pas les syntagmes, complétés dans certains cas par des informations d'ordre sémantique. Pour chaque type de rôle syntaxique sont également indiquées les catégories de mots pouvant remplir le rôle. Enfin, est indiqué le degré de nécessité du syntagme : certains sont toujours obligatoires syntaxiquement ; d'autres peuvent être obligatoires ou facultatifs d'un point de vue syntaxique ou référentiel ; enfin, certains éléments facultatifs peuvent cependant être parfois nécessaire d'un point de vue communicatif.

Rôle syntaxique			**Information sémantique**	**Nature grammaticale**	**Présence obligatoire ou facultative**
Prédicat « PRED »				Syntagme verbal « V », « LV » Syntagme adjectival « ADJ », « LA », « AE » Adverbe « ADV » Syntagme nominal « NOM », « SN(E) », « PRO »	OV
Sujet du prédicat « SUJ »				Clause Syntagme nominal « NOM », « SN(E) », « PRO »	OV
Complément du prédicat	Complément de verbe	Objet « OBJ »		Clause Syntagme nominal « NOM », « SN(E) », « PRO »	OV
		Objet second « OBJ2 »		Clause	OV
		Complément adpositionnel « ADP »	Destinataire « DEST » Oblique « OBQ » Lieu « LOC », « DIR », « SOU »	Syntagme nominal « NOM », « SN(E) », « PRO »	OV
	Complément de nom « CNP »			Clause Syntagme adjectival « ADJ », « AE » Adverbe « ADV » Syntagme nominal « NOM », « SN(E) », « PRO »	OR OC FC
	Complément d'adjectif « CAP »			Clause	OV OC FC

Complément circonstanciel « CIR »			Temps « TPS », « BAL » Lieu « LOC », « DIR », « SOU » Manière « MAN » But « BUT » Cause « CAU » Conséquence « CONS » Instrumental « INST » Condition « COND » Comparaison « COMP » Comitatif « COMI » Agent « AG »	Clause Adverbe « ADV » Syntagme nominal « NOM », « SN(E) », « PRO »	OC FC
Complément interne	Complément d'adjectif « CA » épithète			Clause	OV OC FC
	Complément de nom « CN »			Clause	OR (restrictive) OC (narrative) FC (explicative)
				Syntagme adjectival « ADJ », « AE » Adverbe « ADV » Syntagme nominal « NOM », « SN(E) », « PRO »	OR OC FC
				Possessif « POS »	OR

Mot de liaison « LIA »			Addition « ADD » Opposition / concession « OPP » Hypothèse « HYP » Conséquence « CONS » Alternative « ALTER » Cause, explication « CAU » Modalisateur « MOD »	Conjonction « CON » Adverbe « ADV »	OC FC

Tableau 45. Récapitulatif du codage des syntagmes

1.1.3. Caractéristiques sémantiques et lexicales du texte

Du point de vue du texte dans sa globalité, différentes caractéristiques sémantiques et lexicales peuvent être indiquées. Certains codages présentés ici sont pertinents dans le cadre des textes de notre corpus (suite d'un récit fictif), mais n'ont pas vocation à être pertinents pour tout type de texte. Ils reflètent un certain nombre de choix qui ont dû être effectués par le scripteur lors de sa production. Tout d'abord, certaines caractéristiques concernent le prédicat : temporalité de l'action, plan de l'action, sémantisme du prédicat. Concernant les sujets, nous indiquerons quel type de sujet est mentionné (e.g. personnage principal, autre personnage) et la continuité du sujet par rapport à l'unité précédente. Une indication sémantique supplémentaire concernera les compléments de lieu. Les chaines référentielles seront ensuite codées. Enfin, la proximité sémantique entre les éléments sera établie grâce aux les champs lexicaux des référents et prédicats, qui permettront d'indiquer le degré d'accessibilité de chaque référent, et par un indice de similarité sémantique pour chaque phrase.

1.1.3.1. Prédicats

Temporalité de l'action

Il aurait été possible de situer le temps de l'évènement (T') par rapport au moment de l'énonciation (T0). Cependant, cela ne nous a pas semblé très pertinent dans le cadre de notre étude. En effet, la plupart des récits étaient rédigés au passé, mais non pas parce qu'ils relataient des faits antérieurs au moment de l'énonciation (le récit étant fictif), mais parce qu'il s'agit du temps utilisé généralement dans les récits ; en outre, cela était cohérent avec l'amorce du texte, qui était au passé.

En revanche, il est possible d'utiliser la notion de « temps de référence », proposé par Reichenbach (cité dans Molendijk, 1985). Dès lors, nous avons distingué plusieurs temps possibles, comme résumé dans le Tableau 46 et illustré dans la Figure 54.

temporalité	codage	explication	exemple
Nouveau temps de référence	« r »	Introduit un nouveau temps de référence	Il déshabilla le gardien et *l'habilla de ses vêtements. Il* **mit** le gardien dans son lit. [S15]
Simultané	« s »	Englobe le temps de référence, mais est limité dans le temps	Il ouvrit la porte que le gardien *n'avait pas refermée. Beaucoup* de monde **courait**. [S16]
Général	« g »	Englobe le temps de référence, n'est pas limité dans le temps	Et ils partirent tous les deux en bateau pour la Crète où Dimitri **possédait** une ferme léguée par son parrain [S19]

temporalité	**codage**	**explication**	**exemple**
Passé	« p »	Antérieur au début de l'histoire	Il sortit la lame de rasoir que son ami Rodolphe lui **avait apportée** quelques jours avant au parloir [S13]
Passé dans l'histoire	« p-h »	Antérieur au temps de référence, mais compris dans le temps de l'histoire	*Il enfila ensuite ceux qu'il* **avait prélevés** au gardien [S11]
Irréel	« ir »	Du domaine de l'irréel : postérieur au temps de référence ; but ou devoir. Les prédicats « ir » ne se réalisent pas de manière immédiate au cours de l'histoire.	Ensemble, ils **allaient prouver** *l'innocence de Dimitri [S23]*

Tableau 46. Critères d'assignation de la temporalité de l'action

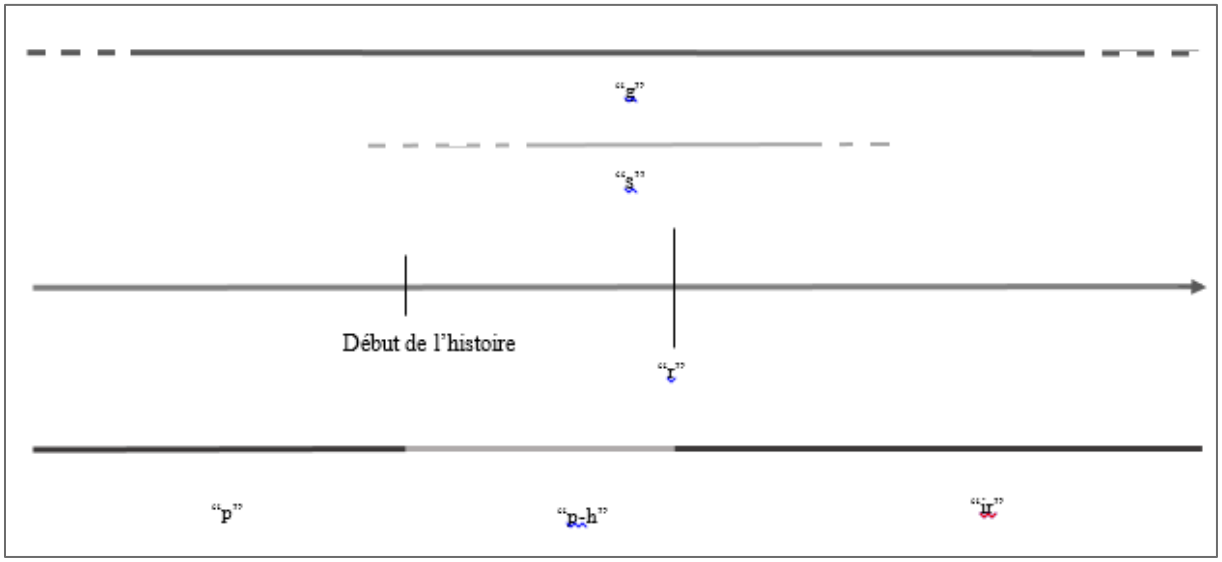


Figure 54. Temporalité de l'action

1er plan et arrière-plan

Les critères pour définir le premier plan et l'arrière-plan ne sont pas toujours clairs, et ne se recoupent pas systématiquement. Pour ce travail, nous avons estimé qu'un prédicat au 1er plan correspondait aux critères suivants :

- exprime une action réellement effectuée ;
- correspond à ce que retiendrait un compte-rendu factuel ;
- est temporellement ordonné : est postérieur au prédicat précédent, et est nécessaire à l'accomplissement du prédicat suivant ;
- est choisi d'être mis au 1er plan par l'auteur (e.g.: par une proposition principale plutôt que subordonnée, et par un verbe au passé simple plutôt qu'à l'imparfait).

Si l'on se rapporte au codage, il s'agit donc de prédicats dont le verbe répond aux critères suivants :

- temporalité : « r » ;
- temps : passé simple (hormis pour les narratifs de narration) ;
- verbe d'action (hormis si le verbe d'état ou de cognition est nécessaire pour l'accomplissement de l'action suivante) ;
- n'appartient pas à du discours direct.

Sémantisme des prédicats

Les verbes ont été classés par classes sémantiques. Ghiglione et al. (1990), dont les travaux sont utilisés par le logiciel Tropes[63], avaient établi une distinction entre verbes « factifs », « statifs » et « déclaratifs ». Des sous-classes avaient été introduites, permettant notamment de subdiviser les « déclaratifs » entre « déclaratifs d'expressions » et « déclaratifs de cognition ». Nous avons choisi de garder cette distinction entre verbe de « parole » et verbe de « cognition ». Nous avons également identifié des distinctions sémantiques importantes par rapport à notre thème. Cela est cohérent avec le fait que Escarabajal et Kekenbosch (1998) avaient montré que les verbes et dérivés verbaux étaient davantage regroupés d'après des critères extrinsèques (liés à la situation) qu'à des propriétés intrinsèques.

La liste ci-dessous indique les catégories sémantiques utilisées avec quelques critères définitoires ; des exemples sont donnés dans le Tableau 47.

- Verbes d'action :
    - Agression : impacte physiquement un autre personnage, de manière négative
    - Interaction : interaction physique avec un autre personnage ou établissement de contact
    - Sens : l'utilisation d'un sens (vue, ouïe, odorat) est bien spécifiée
    - Déplacement : implique un déplacement géographique du sujet
    - Changement de posture
    - Objet : lien physique avec un objet (concret, inanimé), sujet humain
    - Divers (verbes intransitifs, sujets non humains)
- Verbes déclaratifs :
    - Parole : utilise un organe vocal
    - Cognition : l'activité déclarative se situe « dans la tête » du sujet
- Verbes statifs :
    - état
    - pseudo-état : verbes agentifs ou dynamiques utilisés pour décrire un état (Martin, 2007) ; le sujet est généralement non-humain

[63] http://www.tropes.fr

Le Tableau 47 donne quelques exemples pour chaque catégorie sémantique. Une liste plus complète est présentée en Annexe 3.

agression	assassiner ; frapper
interaction	aimer ; inviter ; retrouver
sens	entendre ; épier ; sentir ;
déplacement	entrer ; se faufiler ; venir
changement posture	s'assoir ; tomber
objet	acheter ; ranger ; tenir
action - divers	abandonner ; pleurer ; se changer
parole	chuchoter ; interpeller
cognition	apprendre ; hésiter
état	avoir faim ; être amoureux ;
pseudo-état	enjamber (« un câble qui enjambait le mur d'enceinte »)

Tableau 47. Catégories sémantiques des prédicats - Exemples

Lorsque le prédicat était composé de deux verbes, nous avons utilisé deux colonnes :

- une colonne pour la « modalité » (il s'agit du 1er verbe, généralement conjugué : « pouvoir, vouloir, essayer de… ») ;
- une colonne pour le verbe à l'infinitif.

Une troisième colonne permettait également de préciser lorsqu'il y avait une négation dans le prédicat.

Le sémantisme portait sur l'ensemble du groupe verbal : par exemple, « donner un coup » était considéré comme un prédicat d'agression (ce qui n'est pas le cas de « donner des clés »).

1.1.3.2. Sujets

Type de sujet

Les sujets de chaque prédicat ont été relevés sous leur forme sémantique : lorsqu'un sujet était un pronom ou une ellipse, il a été remplacé par le sujet d'origine. Quatre types de sujets ont été distingués : « Dimitri », « autre personnage », « non personnage », « sujet impersonnel (Ø) ».

Changement ou maintien des sujets

D'un point de vue discursif, il est intéressant de savoir si le sujet est maintenu au long du récit, ou s'il y a des changements de sujet. Il est possible de regarder le changement de sujet à différents niveaux. Ainsi nous avons indiqué si le sujet était changé (« c ») ou maintenu (« m ») dans les cas relevés dans le Tableau 48.

	Phrase précédente	Unité A précédente	Clause précédente
1er sujet de la phrase	1er sujet de la phrase par rapport à la phrase précédente		1er sujet de la phrase par rapport à la clause précédente
1er sujet de l'unité A		1er sujet de l'unité A par rapport à l'Unité A précédente	1er sujet de l'unité A par rapport à la clause précédente
1er sujet de la clause			1er sujet la clause par rapport à la clause précédente

Tableau 48. Unités étudiées concernant la continuité des sujets

Le Tableau 49 analyse des sujets de la phrase (134).

(134) *Il passa alors à l'action.* ~ **Il** avait pris la matraque / que **sa fiancée** avait réussi à lui faire passer lors de sa dernière visite en prison, // et **[ ]** *l'abattit sur le crâne du* gardien. [S22]

		Phrase précédente	Unité A précédente		Clause précédente		
		« **il** (passa alors) »	« **il** (passa alors) »	« **il** (avait pris) »	« **il** (passa alors) »	« **il** (avait pris) »	« que **sa fiancée** (avait réussi) »
1er sujet de la phrase	« **il** (avait pris) »	m			m		
1er sujet de l'unité A	« **il** (avait pris) »		m		m		
	« **[ ]** (l'abattit) »			m			c
1er sujet de la clause	« **il** (avait pris)»				m		
	« **sa fiancée** (avait réussi)»					c	
	« **[ ]** (l'abattit) »						c

Tableau 49. Continuité des sujets - Exemple

1.1.3.3. Lieux

Les indications de lieux (présentes dans des compléments adpositionnels ou des compléments circonstanciels) ont été relevées. Les lieux ont été classés selon deux catégories : « corporel » et « externe ». Le Tableau 50 donne quelques exemples ; une liste plus complète est présentée en Annexe 4.

Corporel	autour de la taille du gardien ; sur sa nuque
Externe	à l'accueil ; dans la rue ; sous son matelas

Tableau 50. Compléments de lieux "externes" et "corporels" - Exemples

1.1.3.4. Chaines référentielles

Lors de la production d'un texte, il est nécessaire d'introduire des référents, puis de les maintenir. Par exemple, dans la phrase suivante, on peut relever trois référents.

(135) **Il** plaqua le **gardien** contre le **mur** [S14]

Le statut informationnel de tous les référents a été codé. La distinction se fait tout d'abord entre référent nouveau et référent ancien ; est considéré comme ancien référent tout référent ayant été mentionné explicitement au préalable. Un ancien référent peut apparaitre essentiellement sous trois formes : une répétition exacte de la mention précédente, une modification avec un nouveau terme ; une modification avec un terme déjà utilisé précédemment. Il peut donc s'agir d'une insertion (« ins »), d'une répétition exacte (« rep »), d'une reprise avec modification sous forme d'insertion d'un nouvel item lexical (« mod ins ») et de réinsertion d'un ancien item lexical (« mod rei »). L'exemple (136) illustre ces quatre possibilités, s'il on s'intéresse au référent « Dimitri ».

(136) **Dimitri** ouvrit la porte. **Il** ne vit personne ; **il** *s'élança dans le couloir*.
« ins » « mod ins » « rep »
Le prisonnier marcha en silence et **il** arriva enfin à la sortie.
« mod ins » « mod rei »

Les participants ayant pour tâche de poursuivre une histoire, certains référents étaient déjà présents dans l'amorce. Les référents de celle-ci ont donc été codés, et ils ont été pris en compte lors du codage des informations quant au numéro d'apparition du référent ou à son statut informationnel.

1.1.3.5. Réseaux sémantiques

Catégorisation des référents en champs lexicaux

Les référents ont été catégorisés de manière sémantique. Tout d'abord, nous avons distingué des catégories permettant de distinguer le type de référent : personnage, partie de l'humain, objet, concept, environnement, temps, quantité.

Nous avons ensuite délimité des champs lexicaux (CL). Ces derniers pouvaient d'une part être liés au fait qu'il s'agit d'un récit d'évasion : on aura donc les CL « Bâtiment », « Prison », « Violence », « Fuite », « Trajet », « Localisation ». De plus, comme des personnages humains sont impliqués dans le récit, il y a des CL liés à l'humain : « Corps », Cognition », « Parole », « Sens », « Sentiment », « Comportement ». Enfin, des CL divers ont été relevés (« ami », « contes », « famille », « animal », « ville », « nature », « repas », « livre », « linge », « divertissement », déménagement », « habitation », « infirmerie », « matière », « sport », « générique », autres)

Le Tableau 51 donne quelques exemples pour chaque catégorie et CL mentionnés. La liste complète des référents du corpus et de leur répartition dans les catégories et CL est présentée en Annexe 5.

PERSONNAGE	Ami	ami ; copain
	Famille	enfant ; mari
	Contes	fée ; princesse
	Prison	bourreau ; prisonnier
	Animal	chat ; rat
OBJET	Corps	chaussette ; veste
	Bâtiment	ampoule ; oreiller
	Fuite	bâillon ; clé
	Violence	arme ; revolver
	Trajet	bateau ; voiture
	Nature	branche pierre
	Repas	bol ; vaisselle
ENVIRONNEMENT	Bâtiment	couloir ; toit
	Prison	grillage ; vestiaire
	Trajet	chemin ; passage
	Nature	arbre ; plage
	Ville	faubourg ; rue
	Localisation	bas ; intérieur
PARTIE DE L'HUMAIN	Corps	bouche ; ventre
	Parole	grognement ; rire
	Cognition	choix ; pensée
	Sens	attention ; sensation
	Sentiment	admiration ; surprise
CONCEPT	Corps	geste ; taille
	Parole	discussion ; terme
	Cognition	notion ; vérité
	Sens	obscurité ; parfum
	Sentiment	meilleur ; soupir
	Prison	emprisonnement ; vol
	Violence	bagarre ; vengeance
	Trajet	demi-tour ; ouverture
	Fuite	alerte ; recherche
	Divertissement	jeu ; fête
	Comportement	habitude ; tâche
TEMPS	Divers	été ; minute
QUANTITE	Divers	filet ; lambeau

Tableau 51. Champs lexicaux des référents - Exemples

Champs lexicaux des prédicats

Nous avons codé l'accessibilité des référents. Pour cela, il était nécessaire de connaitre les prédicats appartenant à un CL tel que défini ci-dessus. Nous avons donc indiqué le CL des verbes qui, pris de manière isolée, étaient typique d'un de ces CL. Par exemple, le verbe « donner » n'est pas lié à un de ces CL ; en revanche, le verbe « courir » est lié au CL du déplacement. La liste des prédicats liés à un CL est donnée en Annexe 6.

Accessibilité des référents

Le codage du statut informationnel avait permis de distinguer les référents « anciens » et « nouveaux », et au sein des anciens référents, le type de reprise effectué (d'un point de vue lexical). Cependant, il se peut qu'un référent soit plus facilement retrouvé s'il appartient au même champ lexical qu'un référent précédemment mentionné. Afin d'avoir une nuance en « nouveau » et « ancien » référent, nous avons donc introduit le statut « accessible », si le référent appartient au champ lexical d'un mot déjà apparu.

Nous avons ainsi indiqué l'accessibilité des référents sur une échelle de 1 (très accessible) à 4 (peu accessible) :

- 1 = ancien référent
- 2 = nouveau référent, d'un CL présent dans la phrase précédente
- 3 = nouveau référent, d'un CL présent dans le texte
- 4 = nouveau référent, nouveau CL

Similarité sémantique

Nous avons utilisé le logiciel d'Analyse sémantique latente (ASL, ou LSA en anglais[64]) afin d'obtenir des indices de similarité sémantique. Pour cela, nous avons eu recours à l'espace sémantique proposé sur le site internet « Français-Total », qui est un espace de dimension 300. Cet espace est basé sur le corpus « Français-Monde » (constitué de 20208 documents issus du journal « Le Monde », soit 8675391 mots au total) et le corpus « Français-Livres » (constitué de 14622 documents, soit 5748581 mots).

Cela nous a permis d'avoir un indice, pour chaque texte, de :

- la similarité sémantique avec les autres textes ;
- la similarité sémantique entre chaque phrase successive ;
- la similarité sémantique entre chaque phrase et tout ce qui précède (amorce comprise).

La similarité est indiquée par un indice allant de -1 (totalement différent) à 1 (totalement identique d'un point de vue sémantique).

[64] http://lsa.colorado.edu/

Soit l'exemple (137). Le signe « # » marque la frontière entre l'amorce et le texte rédigé par le participant. Ici, l'indice de similarité sémantique entre la dernière phrase (en gras) et celle qui la précède (soulignée) est de 0,16 ; en revanche, l'indice de similarité sémantique entre la dernière phrase et l'ensemble du texte qui la précède est de 0,55. Cet exemple montre donc qu'une phrase peut différer notablement de ce qui précède de manière immédiate, mais être tout à fait rattachée à l'espace sémantique de l'ensemble du texte.

(137) *L'*évasion.

Dimitri regardait le ciel étoilé à travers les barreaux. Il devinait les ombres des gardiens dans le mirador. Il entendi*t la porte de sécurité s'ouvrir dans le couloir. Le gardien ouvrit la porte de sa cellule. Il passa alors à l'*action. # Il sortit dans le couloir, regarda à droite et à gauche. Il décida de partir vers la droite. **Il ouvrit la porte de sécurité que le gardi*en n'avait* pas refermée.** [S16]

1.1.4. Informations obtenues grâce aux éléments codés

Grâce aux codages effectués, il est possible de comptabiliser le nombre d'éléments (e.g. nombre de compléments de temps) dans différentes unités (e.g. la clause, la phrase). Il est également possible, pour ces unités, de connaitre la proportion des différents types d'éléments (e.g. proportion de syntagmes obligatoires ou facultatifs par clause ou par phrase).

Quelques pourcentages correspondent à des indices linguistiques fréquemment relevés dans les études. Par exemple, la « densité lexicale » correspond au pourcentage de mots lexicaux (nombre de mots lexicaux divisé par le nombre total de mots), et la « diversité lexicale » correspond au nombre de lemmes divisé par le nombre total de lexèmes. Par ailleurs, le pourcentage de clauses subordonnées correspond à ce qui est parfois appelé « complexité syntaxique ».

En outre, les fréquences de certains éléments ont pu être calculées. La « fréquence » correspond au nombre de fois qu'un élément apparait dans le corpus, divisé par le nombre d'éléments total dans le corpus, le tout multiplié par 1000000. Pour les analyses, nous avons utilisé le log des fréquences. Pour obtenir ces fréquences, nous avons utilisé deux sources :

- le « corpus » : Il s'agit des 30 textes utilisés pour ce travail. Il nous a permis de relever la fréquence des lexèmes, des lemmes et des prédicats.
- la base de données Manulex (Lété, Sprenger-Charolles & Colé, 2004) : Le corpus sur lequel elle s'appuie est constitué de 54 manuels scolaires, du CP au CM2 (soit 1,9 millions de mots).

1.1.5. Résumé

De nombreuses données linguistiques ont donc été codées. Elles sont d'ordres sémantiques, lexicales et syntaxiques. Elles concernent aussi bien le texte dans son ensemble que les unités inférieures (phrases, unités A, clauses).

Pour faire le lien entre données chronométriques et données linguistiques, afin d'étudier en temps réel la micro-planification et la formulation, nous indiquerons d'une part le regroupement en variables sémantiques, syntaxiques et lexicales, et d'autre part les informations disponibles pour chaque unité étudiée.

1.2. Préparation des analyses

1.2.1. Données sélectionnées

Les variables dépendantes sont les données chronométriques. Le choix a été fait de ne pas écarter « arbitrairement » des données, par exemple en déterminant un seuil minimum ou maximum. Certes, cela crée une grande variance dans les données, mais le fait de ne pas avoir moyenné ces données évite d'accorder un poids trop important aux valeurs extrêmes. Cependant, nous avons écarté des données pour lesquelles nous savions que les données chronométriques seraient affectées par des processus rédactionnels différents que dans les autres cas :

- la première phrase du texte ;
- la présence d'une révision ;
- un changement de paragraphe.

Concernant les variables indépendantes nous avons souhaité introduire un maximum de variables linguistiques (sémantiques, syntaxiques, lexicales), caractérisant soit la phrase n, dans le cas des analyses sur les phrases, soit la clause n, dans le cas des analyses sur les clauses. Nous avons cependant vérifié qu'il n'y ait pas de corrélation supérieure à 0,7 : une trop grande dépendance entre les VI fausserait les analyses. Dans le cas des variables catégorielles, nous avons vérifié qu'il y avait au moins une dizaine d'occurrences de chaque modalité dans les données. Au total, 22 ou 23 VI ont été introduites pour chaque modèle. Celle-ci seront précisées avant chaque analyse.

Par ailleurs, même si nous avons souhaité conservé un maximum de phrases malgré leur diversité, nous avons « homogénéisé » le corpus en excluant des cas très particuliers, qui rendaient difficiles la comparaison avec les autres phrases. Nous avons donc écarté les phrases :

- Commençant par une subordonnée (ex : « Quand il arriva dehors, il partit rapidement »), puisqu'il a été montré que la pause initiale était beaucoup plus longue que dans le cas des autres subordonnées, suggérant un traitement différent.
- Contenant une clause enchâssée (ex : « Le message qu'il avait envoyé était bien arrivé »), puisque les clauses ne sont pas écrites successivement, ce qui rendrait confuse la notion de « clause précédente » lors de nos analyses.
- Contenant du discours direct : La présence d'un discours direct marque un changement énonciatif ; en outre, la présence d'un verbe introducteur avant, après ou en incise du discours direct rend également confuse la linéarité de la production des clauses.
- Phrases nominales : Il s'agit souvent de phrases très courtes, ou de phrases isolées de la phrase précédente pour des raisons stylistiques ; en outre, même si elles ont un prédicat, elles ne comportent pas de verbe conjugué. Ces caractéristiques les rendent linguistiquement très différentes des autres phrases, il est donc délicat de les insérer dans le corpus.

Comme nous nous intéressons à l'unité en cours d'écriture et à la dernière clause de la phrase précédente, seules les phrases et les clauses pour lesquelles la phrase n et la clause n-1 ne contenaient aucun des cas ci-dessus ont été gardées pour les analyses.

1.2.2. Groupes crées pour les analyses

Les résultats de la partie sur les unités de segmentation ont suggéré que les clauses ne sont pas toutes planifiées de la même manière : notamment, les clauses subordonnées diffèrent de celles qui ne le sont pas, et les clauses matrices sont précédées d'une pause plus longue que les non-matrices. Nous avons en outre préféré analyser des phrases et des clauses dont la structure est relativement homogène. Par exemple, il est possible qu'une subordonnée qui soit elle-même matrice soit traitée différemment qu'une subordonnée qui clôture l'unité A[65]. Nous avons donc créé des groupes de clauses mutuellement exclusifs, dont les caractéristiques sont données dans le Tableau 52.

Groupe	Critère de sélection	Nombre de données
Phrases	/	N= 246
Clauses autonomes	seule clause de l'unité A	N= 294
Clauses matrices	1ère clause de l'unité A, suivie d'une ou plusieurs subordonnées	N= 114
Clauses subordonnées	2ème et dernière clause de l'unité A	N= 112

Tableau 52. Analyses des phrases et des clauses - Groupes créés

[65] Pour rappel, une unité A est soit composée d'une clause autonome, soit d'une clause matrice avec les subordonnées qui dépendent d'elle.

1.2.3. Analyses utilisées

Des modèles linéaires à effets mixtes ont été utilisés. Les effets aléatoires étaient les participants, et les effets fixes étaient les variables linguistiques. Le modèle retenu (nommé « modèle 2 ») est celui pour lequel le maximum de VI sont entrées simultanément, ce qui permet un maximum de contrôle des variables, puisque cela évite que des effets apparaissent alors qu'ils ne devraient pas (et vice versa). En outre, plus des données linguistiques sont introduites, plus grande est la part de variance expliquée. Nous avons cependant vérifié que le modèle retenu était significativement meilleur que le modèle vide, en utilisant la décroissance de la déviance.

Nous avons également recherché à chaque fois le modèle le « plus parcimonieux » (i.e., le modèle qui ajuste au mieux les données avec le moins de paramètres possible), nommé « modèle 3 ». Dans ce modèle, généralement, il ne reste presque que les VI qui ont un effet significatif. Ce modèle n'a pas été retenu comme modèle final car il expliquait une part moindre de la variance, et certains effets apparaissaient lorsque d'autres variables étaient enlevées du modèle. Cependant, il a été utilisé pour déterminer les variables que nous avons considérées comme ayant un « effet significatif ». En effet, le seuil de significativité de la p-value est établi à 0,05, mais ce seuil peut apparaitre parfois comme arbitraire, surtout dans un cas, comme le nôtre, où nous testons une succession de modèles et que la p-value peut alors légèrement changer d'un modèle à l'autre. Les VI retenues comme ayant un effet sont donc celles pour lesquelles $p < 0,05$ dans le modèle 2 ou $0,05 < p < 0,1$ dans le modèle 2 et $p < 0,05$ dans le modèle 3. Le fait qu'une variable ait un effet tendanciel mais qui se confirme lorsque le modèle devient plus parcimonieux nous a conduit à la considérer comme ayant bien un effet sur la VI qu'il soit pertinent de relever.

Pour obtenir les modèles 2 et 3, des régressions pas-à-pas descendantes ont été effectuées. Avec SPSS, il n'est pas possible de générer automatiquement la procédure « pas-à-pas » pour les modèles linéaires mixtes (contrairement aux modèles de régression sans effets aléatoires). Les différentes étapes ont donc été menées « manuellement », par la procédure suivante :

- seuls les 3 paramètres « constante des effets fixes», « participants », « erreur résiduelle » ont été introduite (modèle 0) ;
- toutes les VI ont été ajoutées (modèle 1) ;
- à chaque étape la VI qui avait le plus petit F a été enlevée, si cette VI avait un $p > 0,1$;
- vers la fin (notamment quand $0,05 < p < 0,1$), plusieurs tests ont été faits pour s'assurer de garder les VI qui donnaient le meilleur modèle.

Parmi ces étapes, il a fallu s'arrêter à différents moments suivant les modèles :

- Pour le modèle 2 : quand la comparaison des log de vraisemblances entre le modèle 0 et le modèle était significative (donc parfois il s'agit du modèle 1, c'est-à-dire avec toutes les VI

entrées, soit un modèle dans lequel certaines VI ont été enlevées). Cela signifie que le modèle 2 est meilleur que le modèle 0, et le maximum de VI ont été insérées.

- Pour le modèle 3 : quand la comparaison des log de vraisemblances entre le modèle et le modèle de l'étape suivante était significative (c'est-à-dire : le fait d'enlever une variable rend le modèle moins bon). Cela signifie que le modèle 3 est meilleur que tout autre modèle qui aurait moins de VI.

2. Temporalité de l'écriture des phrases et des clauses

2.1. Hypothèses

Dans cette partie, nous allons voir dans quelle mesure les variables chronométriques sont affectées par les données linguistiques, pour différentes unités (phrase, clause autonome, clause matrice, clause subordonnée). Ainsi, nous aurons deux types d'information : le moment où les éléments sont planifiés, et le cout de cette planification.

2.1.1. Hypothèses sur le moment de planification

Cette information sera accessible en examinant quelles variables linguistiques affectent quelles variables chronométriques. Notamment, la micro-planification sera manifestée par l'impact de variables sémantiques, alors que la formulation sera manifestée par l'impact de variables syntaxiques et lexicales. Nous verrons notamment si ces catégories de variables affectent plutôt les données chronométriques de l'unité précédente (et sont donc fortement anticipées), la pause initiale de l'unité ou les données chronométriques de l'unité elle-même.

Les principales hypothèses que nous pouvons formuler sont :

- pour la phrase :

La micro-planification commencerait dès la fin de l'écriture de la phrase précédente, alors que la formulation aurait plutôt lieu au début de l'écriture de la phrase. L'hypothèse de travail est donc que la temporalité de l'écriture de la clause précédente sera affectée par les caractéristiques sémantiques de la phrase, alors que la temporalité de la première clause de la phrase sera affectée par les caractéristiques sémantiques, syntaxiques et lexicales.

- pour les clauses :

L'hypothèse principale est que la planification des clauses subordonnées est fortement anticipée lors de l'écriture de la clause matrice, voire même pendant la pause initiale de celle-ci. Les variables linguistiques des clauses subordonnées devraient donc apporter un gain explicatif relativement important à la variance des données chronométriques de la clause précédente. En revanche, un tel effet d'anticipation serait moindre pour les clauses matrices et autonomes, qui sont toujours la première clause de leur unité A.

2.1.2. Hypothèses sur le cout de la planification

La manière dont les variables linguistiques affectent les variables chronométriques, c'est-à-dire le sens de la corrélation (positive ou négative) ou le facteur impliquant une durée plus élevée, peuvent

informer sur les éléments qui entrainent un cout de micro-planification ou de formulation plus élevé ou au contraire moins élevé. Peu d'études ayant été menées dans un contexte similaire au nôtre (production écrite d'un texte spontané, prise en compte simultanée de nombreuses variables), les hypothèses que nous formulons sont basée sur les études qui, comme nous l'avons mentionné dans le contexte théorique, portent parfois sur la production de phrases isolées, ou sur la modalité orale, voire sur une activité de lecture.

Il est donc attendu que certaines caractéristiques rendent la planification plus couteuse (i.e. allongent le temps de pause ou d'exécution) :

- la longueur (Cook et al., 1974), manifestée par exemple par le nombre de référents ou le nombre de séquences ;
- la complexité syntaxique, manifestée par exemple par la profondeur syntaxique, la présence d'un verbe modal ou d'un verbe impersonnel.

À l'inverse, d'autres caractéristiques devraient faciliter ce cout (i.e. diminuer le temps de pause ou d'exécution) :

- la fréquence (Maggio et al., 2012), manifestée par exemple par la fréquence des lemmes et des lexèmes ou la fréquence du prédicat ;
- le lien avec ce qui précède (Pynte, New & Kennedy, 2008b), manifestée par exemple par des référents plus accessibles, le même sujet ou la présence d'une liaison.

2.2. Écriture de la phrase

Le but de cette partie est de voir dans quelle mesure la phrase est planifiée avant ou au début de son écriture.

2.2.1. Variables chronométriques (VD)

Les variables dépendantes des modèles mixtes sont les variables chronométriques liées d'une part à la dernière clause de la phrase précédente, et d'autre part à la première clause de la phrase analysée. Le Tableau 53 indique les 9 VD utilisées dans les analyses portant sur la phrase.

Dernière clause de la phrase n-1	**Phrase n**	**1ère clause de la phrase n**
Écriture interne en ms/seq	Pause initiale de la phrase	Écriture interne en ms/seq
Pauses inter-mots en ms/seq		Pauses inter-mots en ms/seq
Pauses intra-mots en ms/car		Pauses intra-mots en ms/car
Temps d'exécution en ms/car		Temps d'exécution en ms/car

Tableau 53. Variables chronométriques (VD) liées aux phrases

2.2.2. Variables linguistiques (VI)

Quelques corrélations fortes, auxquelles nous pouvions nous attendre, ont été relevées :

- Les variables liées au nombre de clauses, de syntagmes dans les propositions de niveau 1, de référents, d'anciens référents, de nouveaux référents, de mots, de séquences et de caractères, ont toutes des coefficients de corrélations tels que 0,69<r<0,99.
- Pour les variables « Fréquence lexèmes – corpus » et « Fréquence lemmes – corpus », r=0,92.
- Pour les variables « Fréquence lexèmes – Manulex » et « Fréquence lemmes – Manulex », r=0,78.

Les variables « nombre de séquences », « fréquence lexèmes – corpus » et « fréquence lexèmes – Manulex » ont été gardées. En tout, 22 VI, présentées dans le Tableau 54, ont été insérées dans les analyses.

Variables sémantiques	Variables syntaxiques	Variables lexicales
Présence d'un adverbe de liaison	Profondeur syntaxique maximum	Densité lexicale
Fréquence des prédicats	Présence de clauses non finies	Fréquence lexèmes – corpus
Présence d'un prédicat de 1er plan	% clauses subordonnées	Fréquence lexèmes – Manulex
Présence d'un prédicat d'arrière-plan		Nombre de séquences
Type de sujet (absence / Dimitri / autre personnage / non personnage)		
Continuité du sujet par rapport à la clause précédente (changement / maintien)		
Continuité du sujet par rapport au 1er sujet de la phrase précédente (changement / maintien)		
% d'anciens référents		
Difficulté d'accessibilité des référents		
ASL[66] / phrase n-1		
ASL / texte précédent		
% de syntagmes FC au niveau 1		
% de OS (=OR +OV) pour le niveau 1		
Présence d'un complément de temps au niveau 1		
Présence d'un complément de Lieu au niveau 1		

Tableau 54. Variables linguistiques (VI) liées aux phrases

2.2.3. Résultat des modèles linéaires mixtes

Neuf analyses ont été menées, correspondant aux 9 VD. Pour chacune de ces analyses, nous avons le modèle vide, le modèle 2 (meilleur que le modèle 0, avec le maximum de VI insérées) et le modèle 3

[66] indice d'analyse sémantique latente

(modèle le plus parcimonieux). Il serait trop long de présenter successivement tous les résultats de chaque modèle. Cependant, les résultats pertinents sont présentés dans le tableau présent en Annexe 7. Nous y avons indiqué :

- les ICC (indiquant la part de variance expliquée par la variabilité inter-participants) des modèles 0 et 2 ;

- les données liées à la décroissance de la déviance : LL des modèles 0 et 2, nombre de paramètres insérés dans le modèle 2, résultat du test de χ^2 comparant ces deux modèles ;

- le pseudo R^2, c'est-à-dire la part de variance supplémentaire expliquée par les effets fixes (ici, les variables linguistiques) insérées dans le modèle 2 ;

- les VI pour lesquelles la p-value est inférieure à 0,10, ainsi que la p-value et le résultat de l'effet de la VI sur la VD : en gras apparaissent les variables plutôt considérées comme « sémantiques », en souligné les variables « syntaxiques » et en italique les variables « lexicales » ; les VI entre parenthèses ont un effet tendanciel ; seules les VI doublement soulignées seront finalement retenues lors de la discussion, car leur effet est significatif dans le modèle 2 ou tendanciel dans le modèle 2 et significatif dans le modèle 3.

Les résultats que nous conserverons pour la discussion sont synthétisés dans le Tableau 55 : seuls y sont mentionnés la valeur du pseudo R^2 (exprimée en pourcentage) et les VI que nous avons considérées comme ayant un effet significatif, conformément à la méthodologie mentionnée ci-dessus.

	C_{n-1}			C_1		
Pause initiale				10,61%	Présence d'un prédicat de 1er plan	Pst > Abs
					Fréquence des prédicats	-
Pauses intra-mots	13,94%	% de OS au niveau 1	+	17,14%	Présence d'un adverbe de liaison	Pst > Abs
		Difficulté d'accessibilité des référents	+		Type de sujet	Ø > autres
					Fréquence des prédicats	-
					% de syntagmes FC au niveau 1	-
					Fréquence lexèmes – Manulex	-
Exécution	ns			11,17%	% de OS au niveau 1	+
					Fréquence lexèmes – Manulex	-
Pause inter-mots	11,85%	ASL / phrase n-1	+	18,09%	Présence d'un prédicat de 1er plan	Pst > Abs
					Présence d'un complément de lieu au niveau 1	Pst > Abs
					Profondeur syntaxique maximum	+
					Fréquence lexèmes – Manulex	-
Écriture interne[67]	ns			22,65%	Type de sujet	Ø > personnage non Dimitri
					Profondeur syntaxique maximum	+
					Fréquence lexèmes – corpus	-
					Fréquence lexèmes – Manulex	-

Tableau 55. Récapitulatif des effets des VI sur les VD liées aux phrases

[67] Pour rappel, il s'agit du temps mis pour écriture l'unité, pauses et temps d'exécutions compris, après avoir exclu la pause initiale de l'unité.

2.2.4. Discussion

2.2.4.1. Facteurs linguistiques influençant les données chronométriques

Des premières remarques peuvent être émises concernant l'ampleur des effets des VI liées aux phrases sur les variables chronométriques de l'écriture de ces phrases.

Tout d'abord, il était attendu que la phrase soit en partie anticipée avant son écriture, par exemple lors de la fin de l'écriture précédente, mais que seule la micro-planification se déroulerait à ce moment-là. De manière générale, peu d'effets ont été relevés sur les variables chronométriques de C_{n-1} : il n'y a aucun effet significatif sur le temps d'exécution ni sur le temps total d'écriture interne. Les effets sur les pauses intra-mots et inter-mots de la clause précédente ne sont pas très élevés : respectivement, seuls 11,85% et 13,94% de la variance sont expliqués par les VI sélectionnées. Cependant, conformément à notre hypothèse, seules des variables sémantiques ont un effet significatif sur ces VD.

La pause initiale de la phrase, qui rappelons-le est en moyenne très longue (presque 10 secondes), est elle aussi peu expliquée par ces variables linguistiques (10,61%). Il est à noter que seuls des effets liés aux prédicats (présence d'un prédicat de 1er plan, fréquence moyenne des prédicats) ont ici un impact. Cela peut être rapproché de l'étude de Immonen et Mäkisalo (2010), qui montrait que les prédicats étaient précédés d'une pause plus courte pendant l'écriture de la phrase : les auteurs faisaient alors l'hypothèse que le prédicat était planifié avant le début de la phrase, ce qui est corroboré par notre résultat. C'est en outre cohérent avec notre approche linguistique selon laquelle le prédicat est le pivot de la clause, et que c'est à partir de lui que les différents syntagmes s'organisent.

Enfin, les résultats concernant l'écriture de la 1ère clause de la phrase montrent un effet légèrement plus fort de ces VI, puisque toutes les variables chronométriques sont impactées, et que la part de variance expliquée va jusqu'à 22,65% (pour l'écriture interne). Conformément à notre hypothèse, il y a bien des effets sémantiques, syntaxiques et lexicaux. Il est à noter que la variable « Fréquence des lexèmes – Manulex » affecte toutes les VD (pauses intra-mots, exécution, pauses inter-mots, écriture interne).

2.2.4.2. Facteurs linguistiques couteux et facteurs facilitateurs

Nous pouvons à présent regarder plus en détail la manière donc les VI impactent les VD. En effet, certaines caractéristiques entrainent un ralentissement de l'écriture (temps de pause et d'exécution plus long), alors qu'au contraire d'autres l'accélèrent, comme résumé dans le Tableau 56.

+ lent			+ rapide		
Présence d'un complément de lieu au niveau 1	Pst > Abs	Pauses inter-mots C_1	% de syntagmes FC au niveau 1	-	Pauses intra-mots C_1
Présence d'un adverbe de liaison	Pst > Abs	Pauses intra-mots C_1	Fréquence des prédicats	-	Pause initiale phrase Pauses intra-mots C_1
ASL / phrase n-1	+	Pauses inter-mots C_{n-1}	Fréquence lexèmes – Manulex	-	Pauses intra-mots C_1, Exécution C_1, Pauses inter-mots C_1, Écriture interne C_1,
Difficulté d'accessibilité des référents	+	Pauses intra-mots C_{n-1}	Fréquence lexèmes – corpus	-	Écriture interne C_1,
% de OS au niveau 1	+	Pauses intra-mots C_{n-1}, Exécution C_1			
Type de sujet	Ø > autres Ø > personnage non Dimitri	Pauses intra-mots C_1, Écriture interne C_1,			
Présence d'un prédicat de 1er plan	Pst > Abs	Pause initiale phrase, Pause inter-mots C_1			
Profondeur syntaxique maximum	+	Pauses inter-mots C_1, Écriture interne C_1,			

Tableau 56. Sens des effets des facteurs linguistiques sur les données chronométriques liées à la phrase

Tout d'abord, un résultat peut paraitre surprenant : plus il y a de lien entre une phrase et la précédente, plus les pauses sont longues. En effet, la variable « Analyse sémantique latente (ASL) par rapport à la phrase précédente » est positivement corrélée à la longueur des pauses inter-mots de la clause précédente, et la présence d'un adverbe de liaison conduit à des pauses intra-mots de la 1ère clause de la phrase plus longues. Nous aurions pu au contraire supposer que le fait de produire une phrase fortement liée à la première ait un effet facilitateur, puisqu'il serait moins difficile d'accéder aux référents, et la phrase pourrait déjà être anticipée en amont. Ce résultat rappelle plutôt la difficulté à maintenir une cohérence entre les phrases. En effet, la présence de référents similaires conduit à s'interroger sur la meilleure forme lexicale à utiliser, et la présence d'un adverbe de liaison nécessite de s'assurer que le lien logique est effectivement présent. Si le fait d'avoir un lien sémantique fort avec la phrase précédent a sans doute un effet facilitateur en lecture, comme en témoignent des durées liées aux mouvements oculaires plus courtes (Pynte, New & Kennedy, 2008a), ce fort degré de liaison entraine vraisemblablement un cout en production.

Les variables chronométriques sont également plus longues lorsque des éléments plus « difficiles » à produire sont insérés. Par exemple, si les référents sont en moyenne plus difficiles d'accès (car étant

nouveaux et n'appartenant pas nécessairement à un champ lexical utilisé peu de temps avant), les pauses intra-mots sont plus longues dans la clause précédente. Cette « difficulté » peut être également d'ordre syntaxique, au niveau de la 1ère clause de la phrase : par exemple, plus la profondeur syntaxique est élevée (c'est-à-dire, plus les syntagmes sont enchâssés), plus les pauses inter-mots sont longues, ce qui se reflète même au niveau du temps total d'écriture interne. Cela signifie également que dès la 1ère clause de la phrase, le scripteur a un schéma général de la structure de sa phrase, même si cela ne signifie pas que tous les syntagmes soient déjà planifiés. Enfin, une difficulté d'ordre sémantique, mais qui affecte aussi dans une certaine mesure la syntaxe, concerne le type de sujet. Il apparait que les phrases sont plus longues à produire lorsque le sujet est « impersonnel » (c'est-à-dire, lorsque le sujet apparent est un « il » qui n'a pas de référent).

Un résultat robuste concerne l'impact de la fréquence des lexèmes, puisque cette variable lexicale affecte toutes les variables chronométriques de la 1ère clause : plus les lexèmes utilisés sont fréquents, plus l'écriture est rapide, puisqu'il y a une corrélation négative entre ces variables. La fréquence des prédicats a également un effet facilitateur, durant la pause initiale de la phrase ainsi que durant les pauses intra-mots de la 1ère clause. Le fait d'utiliser des prédicats fréquents renvoie notamment à la notion de script (Abelson, 1981). Certaines actions sont en effet davantage attendues dans un récit d'évasion. Par conséquent, même si ce prédicat n'a pas nécessairement été utilisé dans ce texte, son activation est plus aisée car il est attendu pour ce type de texte, et sans doute facilement accessible.

L'effet des variables chronométriques donne en outre des informations sur la planification des éléments obligatoires et facultatifs. En effet, il apparait que le facteur lié au pourcentage de syntagmes obligatoires a un effet dès la clause précédente, contrairement au facteur lié au pourcentage de syntagmes facultatifs. Le fait d'avoir beaucoup d'éléments obligatoires allonge la durée des pauses ou le temps d'exécution, alors que les pauses intra-mots de la 1ère clause sont plus courtes lorsqu'il y a davantage de syntagmes facultatifs. Pour rappel, seuls les syntagmes « de niveau 1 » sont ici pris en compte. Cela signifie que, dès la fin de la phrase précédente, le scripteur a une idée des principaux constituants de la phrase suivante, notamment ceux qui seront nécessaires pour saturer la valence du prédicat ou s'assurer que le référent sera bien déterminé. En revanche, il pourra attendre le début de l'écriture de la phrase pour réfléchir à l'insertion d'éléments qui sont facultatifs d'un point de vue communicatif. En outre, nous pouvons considérer que les prédicats de 1er plan sont eux aussi, mais à un niveau discursif, « obligatoires ». Il s'agit en effet des prédicats qui servent à faire avancer l'histoire, et dont l'omission pourrait rendre l'ensemble moins cohérent, donc moins compréhensible. Or, leur présence conduit à des pauses plus longues en début de phrase ou entre les mots de la 1ère clause.

Enfin, il peut être noté que la présence d'un complément de lieu conduit à un allongement des pauses inter-mots de la 1ère clause.

2.3. Écriture des clauses

Cette partie a pour but de mieux comprendre la temporalité de l'écriture des clauses, en examinant à quel moment les éléments sémantiques, syntaxiques et lexicaux sont planifiés. La planification des clauses varie certainement suivant le lien de dépendance entretenu avec les clauses environnantes. Nous avons donc conduit les analyses sur trois types de clauses : les clauses matrices (1ères clauses de leur unité A), les clauses subordonnées (2èmes et dernières de l'unité A) et les clauses autonomes (uniques clauses de leur unité A).

2.3.1. Présentation de l'analyse

2.3.1.1. Variables chronométriques (VD)

Les variables dépendantes des modèles mixtes sont les variables chronométriques de la clause étudiée et de la clause précédente. Le Tableau 53 indique les 10 VD utilisées dans les analyses portant sur chaque groupe de clauses.

Clause n-1	Clause n
Pause initiale (en ms)	Pause initiale (en ms)
Écriture interne (en ms/seq)	Écriture interne (en ms/seq)
Pauses inter-mots (en ms/seq)	Pauses inter-mots (en ms/seq)
Pauses intra-mots (en ms/car)	Pauses intra-mots (en ms/car)
Temps d'exécution (en ms/car)	Temps d'exécution (en ms/car)

Tableau 57. Variables chronométriques (VD) liées aux clauses

2.3.1.2. Variables linguistiques (VI)

Les VI sont les variables linguistiques caractérisant les clauses étudiées : les éléments comptabilisés ou moyennés le sont donc au niveau de la clause ; de même, les syntagmes étudiés (prédicats, complément de temps, etc.) sont uniquement ceux appartenant à la clause. En tout, 29 variables linguistiques caractérisent chaque clause. Cependant, certaines VI seront enlevées ou adaptées pour chaque groupe d'analyse, afin qu'aucune corrélation ne soit trop forte (il faut que $r<0,70$) et qu'il y ait un nombre suffisant de données par modalité dans le cas des variables nominales.

Ainsi, dans les 3 groupes de variables, il y avait une corrélation forte entre le nombre de mots, de séquences et de caractères : à chaque fois, seul le nombre de séquences a été conservé. Par ailleurs, nous avons procédé à quelques regroupements concernant le sémantisme des prédicats : les 4 modalités identifiées sont « actions / cognition, parole, sens / déplacement / état, pseudo-état ».

Variables sémantiques	Variables syntaxiques	Variables lexicales
% de syntagmes OS (OV + OR)	Profondeur syntaxique maximum dans la clause	Nombre de séquences
% de syntagmes OC	Verbe modal (présence de + d'un verbe)	Densité lexicale
présence de syntagmes FC		Fréquence des lexèmes – corpus
Nb de complément de lieu		Fréquence des lemmes – corpus
Type de complément de lieu (externe / corporel)		Fréquence des lemmes – Manulex
Nb de complément de temps		Fréquence des lexèmes – Manulex
Type de complément de temps (balise / autre)		% de référents avec insertion de l'item lexical
Présence d'une liaison		% de référents avec réinsertion de l'item lexical
Plan du prédicat (1er plan / arrière-plan)		% de référents avec répétition de l'item lexical
Temporalité du prédicat (actions / cognition, parole, sens / déplacement / état, pseudo-état)		Nature grammaticale du sujet (aucun sujet / ellipse / nom propre / syntagme nominal / pronom)
Sémantique du prédicat		
Fréquence du prédicat		
Type de sujet (Dimitri / autre personnage / rien / autre)		
Continuité du sujet par rapport à la clause précédente		
Nb de référents (total)		
Difficulté d'accessibilité des référents		
% de référent « personnage »		
ASL / clause n-1		

Tableau 58. Variables linguistiques (VI) liées aux clauses

2.3.1.3. Présentation des résultats et de la discussion

Les analyses ont été menées pour chaque type de clause : clause matrice, clause subordonnée, clause autonome. Les résultats sont tout d'abord indiqués pour chacun de ces groupes. Le tableau récapitulatif donnant les détails des résultats des modèles est fourni en annexe ; il se présente de la même manière

que celui des phrases. Pour chaque type de clause, nous présentons un tableau récapitulatif des VI ayant un effet sur la VD. Une première analyse de ces résultats est proposée à chaque fois, afin de faire ressortir pour chaque type de clauses les VI ayant un effet significatif, ce qui renseigne sur le moment où ces éléments sont planifiés. La discussion concernant le type d'effet (couteux ou facilitateur) se fera à l'issue de la présentation des résultats des trois types de clauses, et portera sur l'ensemble des trois groupes.

2.3.2. Résultats pour chaque type de clause

2.3.2.1. Clauses matrices

Variables linguistiques (VI) retenues

Le nombre de référents est fortement lié au nombre de séquences (r=0,73), cette variable n'a donc pas été conservée.

Par ailleurs, quelques variables catégorielles ont été réorganisées ou remplacent des variables continues afin d'assurer un nombre minimum de données par modalité :

- Une variable « présence de complément de lieu » remplace la variable continue « nombre de compléments de lieu », et le type de complément de lieu n'est pas précisé.
- Une seule variable concerne les compléments de temps, avec les 3 modalités « aucun / balise / autre ».

Les temporalités « p » (passé) et « p-h » (passé dans l'histoire) ont été regroupées en une seule modalité.

Au total, 24 VI ont donc été retenues pour les analyses.

Résultats : Variables ayant un effet significatif

Les résultats sont présentés dans l'Annexe 8. La présentation des résultats est la même que pour les phrases : y sont indiqués les ICC des modèles 0 et 2 ; les données liées à la décroissance de la déviance des modèles 0 et 2 ; le pseudo R^2 lié aux variables linguistiques insérées dans le modèle 2 ; les VI pour lesquelles la p-value est inférieure à 0,10, ainsi que la p-value et le résultat de l'effet de la VI sur la VD. Les variables retenues, présentées dans le Tableau 59, sont en soulignage double dans le tableau des résultats.

Le Tableau 59 récapitule les VI ayant un effet significatif sur la VD ainsi que la part de variance expliquée par l'introduction de ces VI. Tout comme dans le tableau des résultats en annexe, nous avons distingué les facteurs en fonction de leur catégorie linguistique : les facteurs en gras concernent la sémantique, les facteurs soulignés concernent des choix syntaxiques, les facteurs en italique concernent des choix lexicaux.

	C_{n-1}		C_1	
Pause initiale	39%	**Présence de syntagme FC**	44%	**Présence de complément de lieu**
		Type de complément de temps		**Présence d'une liaison**
		Temporalité du prédicat		Profondeur syntaxique maximum
		Type de sujet		Densité lexicale
		Difficulté d'accessibilité des référents		Fréquence des lexèmes –Manulex
		% de référent « personnage »		
		% de référents avec insertion de *l'item lexical*		
		% de référents avec réinsertion de *l'item lexical*		
		% de référents avec répétition de *l'item* lexical		
		Densité lexicale		
		Fréquence des lexèmes – corpus		
Pauses intra –mots	19%	**Continuité du sujet par rapport à la clause précédente**	33%	**Sémantique du prédicat**
		Nature grammaticale du sujet		**Continuité du sujet par rapport à la clause précédente**
		Nombre de séquences		Nature grammaticale du sujet
				Profondeur syntaxique maximum
				% de référents avec réinsertion de *l'item lexical*
				Nombre de séquences
Exécution	31%	**Présence d'une liaison**	39%	**Présence de complément de lieu**
		Temporalité du prédicat		**Plan du prédicat**
		Nature grammaticale du sujet		**Type de sujet**
		% de référents avec réinsertion de *l'item lexical*		**Continuité du sujet par rapport à la clause précédente**
		Densité lexicale		Présence d'un verbe modal
		Fréquence des lemmes – Manulex		(% de référents avec répétition de *l'item lexical)*
				Fréquence des lemmes – corpus
Pause inter-mots	34%	**Présence de complément de lieu**	41%	**Présence de syntagme FC**
		Continuité du sujet par rapport à la clause précédente		**Type de complément de temps**
		Fréquence des lemmes – corpus		**Type de sujet**
		Fréquence des lemmes – Manulex		% de référents avec réinsertion de *l'item lexical*
				% de référents avec répétition de *l'item lexical*
				Densité lexicale

Écriture interne	13%	**Type de complément de temps**	62%	**Présence de syntagme FC**
		Densité lexicale		**Présence de complément de lieu**
				Type de complément de temps
				Sémantique du prédicat
				Type de sujet
				Nature grammaticale du sujet
				Profondeur syntaxique maximum
				% de référents avec insertion de *l'item lexical*
				Nombre de séquences

Tableau 59. Récapitulatif des effets des VI sur les VD liées aux clauses matrices

Globalement, nous pouvons observer qu'une part assez importante de la variance peut être expliquée par les VI introduites. Ainsi, il y 44% de gain explicatif grâce aux VI concernant la pause initiale de la clause, et 39% pour la pause initiale de la clause précédente. Ce résultat peut paraitre surprenant puisque pour ces analyses la clause matrice est toujours la 1ère clause de l'unité A ; elle n'a donc pas de lien de dépendance syntaxique avec la clause précédente. Nous nous attendions donc à ce que les données chronométriques de la clause précédente soient peu affectées par les VI de la clause matrice. Cependant, nous pouvons voir que de manière globale la durée d'écriture interne de la clause précédente est peu affectée (13%) par ces VI. À l'inverse, le gain explicatif est de 62% pour l'écriture interne de la clause en cours d'écriture ; les VI affectant la variable « écriture interne » affectent généralement aussi une des variables de cette écriture (pauses inter-mots, intra-mots ou exécution).

Les effets significatifs ne permettent de dégager un pattern du type : micro-planification pendant la clause précédente, micro-planification et formulation pendant la clause en cours d'écriture, comme c'était le cas pour les phrases. Nous observons que les variables syntaxiques (en italique) n'affectent que la clause en cours d'écriture, à partir de la pause initiale de celle-ci, alors que les VD liées à la clause précédente sont expliquées par des variables sémantiques et lexicales. Durant la pause initiale de C_{n-1}, les VI ayant un effet significatif sont relatives à différents aspects sémantiques : prédicat, sujet, complément de temps, référents. L'écriture elle-même de la clause est davantage affectée par des VI liées au lien entre les clauses : continuité du sujet par rapport à la clause précédente, présence d'une liaison. Cela souligne le fait que, même si les clauses n'entretiennent pas de lien de dépendance syntaxique fort, elles entretiennent nécessairement des liens d'ordre sémantique, puisque leur succession permet de créer un texte.

2.3.2.2. Clauses subordonnées

Variables linguistiques (VI) retenues

Quelques corrélations fortes ont été repérées :

- La fréquence des lemmes et celle des lexèmes sont fortement corrélées, que ce soit dans la base « corpus » (r=0,90) ou « Manulex » (r=0,74) ; seules les fréquences de lexèmes ont été conservées.

Par ailleurs, quelques variables catégorielles ont été réorganisées ou remplacent des variables continues afin d'assurer un nombre minimum de données par modalité :

- Une variable « présence de complément de lieu » remplace la variable continue « nombre de compléments de lieu », et le type de complément de lieu n'est pas précisé.
- Une variable « présence de complément de temps » remplace la variable continue « nombre de compléments de temps », et le type de complément de temps n'est pas précisé.

Au total, 22 VI ont donc été retenues pour les analyses.

Résultats

Le détail des résultats est présenté dans l'Annexe 9. Le Tableau 60 récapitule les résultats, en indiquant le gain explicatif dû aux variables linguistiques (exprimé en pourcentages) ainsi que les variables linguistiques ayant un effet significatif sur la VD concernée.

<table>
<tr><th></th><th colspan="2">C_{n-1}</th><th colspan="2">C_n</th></tr>
<tr><td rowspan="3">Pause initiale</td><td rowspan="3">24%</td><td>Nature grammaticale du sujet</td><td rowspan="3">42%</td><td>Sémantique du prédicat</td></tr>
<tr><td>% de référents avec répétition de l'item lexical</td><td>(% de référents avec réinsertion de l'item lexical)</td></tr>
<tr><td>Nombre de séquences</td></tr>
<tr><td rowspan="6">Pauses intra–mots</td><td rowspan="6">26%</td><td>Présence de syntagme FC</td><td rowspan="6">31 %</td><td>Présence de syntagme FC</td></tr>
<tr><td>Présence d'un complément de lieu</td><td>Temporalité du prédicat</td></tr>
<tr><td>Plan du prédicat</td><td rowspan="4">% de référents avec réinsertion de l'item lexical</td></tr>
<tr><td>Sémantique du prédicat</td></tr>
<tr><td>Présence d'un verbe modal</td></tr>
<tr><td>Nombre de référents</td></tr>
<tr><td rowspan="5">Exécution</td><td rowspan="5">54%</td><td>Sémantique du prédicat</td><td rowspan="5">44%</td><td>Continuité du sujet par rapport à la clause précédente</td></tr>
<tr><td>Type de sujet</td><td rowspan="4">Nature grammaticale du sujet</td></tr>
<tr><td>Présence d'un verbe modal</td></tr>
<tr><td>Fréquence du prédicat</td></tr>
<tr><td>% de référents avec répétition de l'item lexical</td></tr>
<tr><td rowspan="4">Pause inter-mots</td><td rowspan="4">31%</td><td rowspan="4">Profondeur syntaxique maximum</td><td rowspan="4">50%</td><td>Présence d'un complément de lieu</td></tr>
<tr><td>Plan du prédicat</td></tr>
<tr><td>Temporalité du prédicat</td></tr>
<tr><td>Nombre de séquences</td></tr>
<tr><td rowspan="2">Écriture interne</td><td rowspan="2">9 %</td><td>Nombre de référents</td><td rowspan="2">48 %</td><td>Temporalité du prédicat</td></tr>
<tr><td>% de référent « personnage »</td><td>Densité lexicale</td></tr>
</table>

Tableau 60 Récapitulatif des effets des VI sur les VD liées aux clauses subordonnées

Le gain explicatif s'élève entre 30 et 50% pour la clause en cours d'écriture. Il est en revanche plus faible pour la clause précédente : généralement un peu moins de 30%, avec seulement 9% pour l'écriture interne. Il faut cependant noter que le gain explicatif le plus élevé se trouve bien pour une VD de C_{n-1} : 54% pour l'exécution. Il se pourrait alors la planification de la clause subordonnée soit principalement anticipée pendant l'écriture de la clause matrice. En outre, contrairement aux résultats observés pour la clause matrice, les VI syntaxiques (présence d'un verbe modal, profondeur syntaxique maximum) affectent l'écriture de la clause précédente.

2.3.2.3. Clauses autonomes

Variables linguistiques (VI) retenues

Quelques corrélations fortes ont été repérées :

- La fréquence des lemmes et celle des lexèmes sont fortement corrélées, que ce soit dans la base « corpus » (r=0,9) ou « Manulex » (r=0,76) ; seules les fréquences de lexèmes ont été conservées.

Par ailleurs, quelques variables catégorielles ont été réorganisées ou remplacent des variables continues afin d'assurer un nombre minimum de données par modalité :

- Une seule variable concerne les compléments de lieu, avec les 3 modalités « aucun / externe / corporel ».
- Une seule variable concerne les compléments de temps, avec les 3 modalités « aucun / balise / autre ».
- Les temporalités « p » (passé) et « p-h » (passé dans l'histoire) ont été regroupées en une seule modalité.

Au total, 23 VI ont donc été retenues pour les analyses.

Résultats

Le détail des résultats est présenté dans l'Annexe 10. Le Tableau 61 récapitule les résultats, en indiquant le gain explicatif dû aux variables linguistiques (exprimé en pourcentages) ainsi que les variables linguistiques ayant un effet significatif sur la VD concernée.

	C_{n-1}		C_n	
Pause initiale	19%	**Fréquence des prédicats**	27%	**Présence de syntagmes FC**
				Présence d'un adverbe de liaison
				Plan du prédicat
				Fréquence des lemmes – corpus
Pauses intra – mots	11%	Nature grammaticale du sujet	13%	**Temporalité du prédicat**
Exécution	ns		13%	**Plan du prédicat**
				% de référents avec répétition de l'item lexical
				Fréquence des lexèmes – Manulex
				Fréquence des lexèmes – corpus
Pause inter-mots	ns		21%	**Sémantique du prédicat**
				Nature grammaticale du sujet
				Profondeur syntaxique maximum
Écriture interne	9%	**Présence d'un adverbe de liaison**	22%	**Sémantique du prédicat**
				% de référents avec répétition de l'item lexical
				Densité lexicale
				Fréquence des lemmes – Manulex

Tableau 61. Récapitulatif des effets des VI sur les VD liées aux clauses autonomes

La variance des données chronométriques concernant les clauses autonomes est globalement peu expliquée par les variables linguistiques introduites. Le gain explicatif le plus élevé concerne la pause initiale de la clause (27%). En revanche, les variables de l'écriture interne de C_{n-1} sont peu, voire pas du tout affectées par les variables linguistiques de C_n. Il semblerait donc que, dans le cas de clauses autonomes (c'est-à-dire les cas où la clause est seule dans l'unité A), il y ait peu d'anticipation. La seule VI ayant un effet significatif sur la pause initiale de la clause précédente est la fréquence du prédicat. Globalement, les VI sémantiques retenues ici concernent essentiellement le prédicat : effet de la fréquence du prédicat sur la pause initiale de C_{n-1}, plan du prédicat (pause initiale et temps d'exécution de de Cn), sémantique du prédicat (pauses inter-mots, écriture interne) ; on trouve également des effets concernant la liaison, que ce soit sur la clause précédente ou la pause initiale de la clause en cours. Les variables lexicales affectent essentiellement C_n, même si le choix de la nature grammaticale du sujet a un effet dès l'écriture de la clause précédente.

2.3.3. Résultats concernant le cout des VI

Afin de discuter du cout de traitement suivant les modalités des variables linguistiques, nous les avons résumées dans le Tableau 62. Le signe « > » signifie que la présence de cette modalité entraine une écriture plus lente (durée des pauses ou des temps d'exécution plus longs). De même, le signe « + » indique que, dans le cas d'une variable continue, l'augmentation de la valeur entraine une durée plus longue. Les modalités les plus « élevées » sont donc les plus couteuses.

	Clauses matrices		**Clauses subordonnées**		**Clauses autonomes**	
Présence de syntagme FC	Pst > Abs Pst > Abs Pst > Abs	Pause initiale C_{n-1} Pause inter –mots Cn Écriture interne Cn	Pst < Abs Pst < Abs	Pauses intra –mots C_{n-1} Pauses intra –mots Cn	Pst < Abs Pst < Abs	Pause initiale Cn
Type de complément de temps	balise < aucun, autre autre > balise aucun > balise aucun > balise	Pause initiale C_{n-1} Écriture interne C_{n-1} Pause inter –mots Cn Écriture interne Cn				
Temporalité du prédicat	simultané < général (général > passé, passé dans histoire)	Pause initiale C_{n-1} Exécution C_{n-1}	ns ns passé > irréel	Pauses intra –mots Cn Pause inter –mots Cn Écriture interne Cn	ns	Pauses intra –mots Cn
Sémantique du prédicat	état > déplacement état > déplacement	Pauses intra –mots Cn Écriture interne Cn	déplacement, état < autres action < déplacement, état action < déplacement, état	Pauses intra –mots C_{n-1} Exécution Cn-1 Pause initiale Cn	état > action état < cognition, parole, sens, déplacement	Pause inter –mots Cn Écriture interne Cn
Type de sujet	autre personnage > Dimitri, non personnage Dimitri > autre personnage (Dimitri > autre personnage) Dimitri > autre personnage, non personnage	Pause initiale C_{n-1} Exécution Cn Pause inter –mots Cn Écriture interne Cn	non personnage < Dimitri, autre personnage	Exécution C_{n-1}		
Difficulté d'accessibilité des référents	-	Pause initiale C_{n-1}				
% de référent « personnage »	-	Pause initiale C_{n-1}	-	Écriture interne C_{n-1}		

Continuité du sujet par rapport à la clause précédente	maintien > changement maintien > changement maintien < changement maintien < changement	Pauses intra –mots Cn-1 Pause inter –mots Cn-1 Pauses intra –mots Cn Exécution Cn	maintien > changement	Exécution Cn		
Présence d'une liaison	Pst < Abs Pst > Abs	Exécution C_{n-1} Pause initiale Cn			Pst > Abs Pst < Abs	Écriture interne C_{n-1} Pause initiale Cn
Présence de complément de lieu	Pst < Abs Pst > Abs Pst > Abs Pst > Abs	Pause inter–mots C_{n-1} Pause initiale Cn Exécution Cn Écriture interne Cn	Pst > Abs	Pauses intra –mots C_{n-1} Pause inter –mots Cn		
Plan du prédicat	1er > arrière	Exécution Cn	1er > arrière 1er < arrière	Pauses intra –mots C_{n-1} Pause inter –mots Cn	1er > arrière	Pause initiale Cn Exécution Cn
Nombre de référents			-	Pauses intra –mots C_{n-1} Écriture interne C_{n-1}		
Fréquence du prédicat			-	Exécution Cn-1	-	Pause initiale C_{n-1}
Profondeur syntaxique maximum	- + +	Pause initiale Cn Pauses intra –mots Cn Écriture interne Cn	+	Pause inter –mots C_{n-1}	+	Pause inter –mots Cn
Présence d'un verbe modal	Pst > Abs	Exécution Cn	Pst > Abs	Pauses intra –mots C_{n-1} Exécution C_{n-1}		
% de référents avec *insertion de l'item* lexical	- +	Pause initiale C_{n-1} Écriture interne Cn				
% de référents avec *réinsertion de l'item* lexical	- + - -	Pause initiale C_{n-1} Exécution C_{n-1} Pauses intra –mots Cn Pause inter –mots Cn	-	Pause initiale Cn Pauses intra –mots Cn		

% de référents avec *répétition de l'item* lexical	-	Pause initiale C_{n-1} Exécution Cn Pause inter –mots Cn	-	Pause initiale C_{n-1} Exécution C_{n-1}	+	Exécution Cn Écriture interne Cn
Densité lexicale	- - - + +	Pause initiale C_{n-1} Exécution $_{Cn-1}$ Écriture interne C_{n-1} Pause initiale Cn Pause inter-mots Cn	+	Écriture interne Cn	+	Écriture interne Cn
Fréquence des lexèmes – corpus	+	Pause initiale C_{n-1}				
Fréquence des lemmes – corpus	+ -	Pause inter-mots C_{n-1} Exécution Cn			- +	Pause initiale Cn Exécution Cn
Fréquence des lexèmes – Manulex	+	Pause initiale Cn				
Fréquence des lemmes – Manulex	- -	Exécution C_{n-1} Pause inter-mots C_{n-1}			-	Exécution Cn Écriture interne Cn
Nombre de séquences	-	Pauses intra - mots C_{n-1} Pauses intra - mots Cn Écriture interne Cn	+	Pause initiale Cn Pause inter - mots Cn		
Nature grammaticale du sujet	(syntagme nominal > ellipse) ellipse > nom propre, pronom ellipse < nom propre, pronom ns	Pauses intra - mots C_{n-1} Exécution C_{n-1} Pauses intra - mots Cn Écriture interne C_n	Pronom > syntagme nominal syntagme nominal > pronom	Pause initiale C_{n-1} Exécution Cn	Ellipse > pronom ellipse > nom propre, pronom	Pauses intra - mots C_{n-1} Pause inter - mots Cn

Tableau 62. Variables linguistiques - effets observés sur les clauses

Grâce à ce tableau, nous pouvons identifier des caractéristiques des clauses qui rendent leur traitement plus couteux ou au contraire le facilite. Certains résultats corroborent les hypothèses formulées ; ils confirment parfois ce qui avait été observé au niveau de la phrase. D'autres résultats sont plus difficiles à interpréter ; leur effet dépend par exemple de la localisation de l'effet (C_{n-1} ou C_n) ou des interactions possibles avec d'autres VI.

Nous présenterons les résultats par catégories de VI. Nous distinguerons tout d'abord ce qui a trait à la constitution du récit (e.g. type de prédicat, de complément, de référents introduits) : un certain nombre de modalités de cette catégorie sont spécifiques à notre étude, les résultats ne seront donc pas forcément généralisables à d'autres types de texte. Nous mentionnerons ensuite les variables relevant de la continuité discursive (liaison entre les clauses, continuité du sujet, accessibilité des référents, continuité référentielle). Nous aborderons ensuite les VI relevant des choix syntaxiques (présence d'un verbe modal, profondeur syntaxique), et évoquerons enfin les caractéristiques liées aux choix lexicaux (nombre de séquences, fréquence des lemmes et lexèmes, nature grammaticale du sujet).

Beaucoup de variables linguistiques liées au récit ont été codées. Dans certains cas, les modalités relevées pour chaque variable dépendent du type de récit lui-même (récit d'évasion), et semblaient les plus pertinentes pour rendre compte de nos données. Ces données concernent les éléments constitutifs d'une clause : sujet, prédicat, compléments de temps et de lieu, référents, constituants facultatifs. Concernant le type de sujet, la présence du personnage « Dimitri » entraine généralement une production[68] plus lente que lorsque le sujet est un autre personnage, hormis dans le cas de la clause précédant la clause matrice. En outre, lorsque la modalité « non personnage » a un effet significatif, c'est toujours dans le sens d'une plus grande rapidité. Le prédicat a été caractérisé par sa fréquence dans le corpus, le plan, la temporalité et la sémantique. Il apparait que les prédicats les plus fréquents entrainent une diminution de la durée des pauses ou du temps d'exécution, cet effet portant sur les clauses précédentes. Une augmentation de la durée est à l'inverse observée lorsqu'il s'agit d'un prédicat de 1er plan, sauf dans le cas de la subordonnée ; dans ce cas, on note qu'un effet d'allongement est observé sur la clause précédente, ce qui signifie que le plan du prédicat est déjà anticipé et couteux pendant C_{n-1} puis plus rapide pendant la production de C_n, alors que dans les autres cas (clauses matrices et autonomes) l'effet, couteux, n'apparait que sur C_n. Relativement peu de résultats émergent concernant la temporalité du prédicat lorsque l'on observe les différentes modalités ; on note tout de même que la temporalité « général » est plus lente pour la production de clauses matrices, cet effet portant sur la clause précédente. Les résultats sont également confus pour l'aspect sémantique du prédicat : lorsque la modalité « cognition, parole, sens » est significative, elle implique un allongement de la durée par

[68] Pour cette partie, nous utiliserons le terme « production » par commodité : il s'agit des pauses et/ou temps d'exécution. Ainsi, une « production lente » signifie qu'au moins une des variables chronométriques de la clause est plus lente que dans les autres modalités.

rapport à la modalité « état » ; à l'inverse, la catégorie « action » est généralement plus rapide que les verbes concernant l'état et le déplacement. Lorsque ces deux dernières catégories sémantiques se distinguent, ce n'est pas dans le même sens : la production est plus rapide pour les prédicats d'« état » pour les clauses matrices, mais plus lente pour les clauses autonomes. Les aspects sémantiques concernant les référents affectent les clauses précédentes : les durées sont plus courtes s'il y a davantage de référents « personnages », et si le nombre de référents dans la clause est plus élevé. Enfin, la présence de syntagmes facultatifs d'un point de vue communicatif affecte différemment les types de clauses : cette présence entraine un cout pour les clauses matrices, mais une facilitation pour les clauses subordonnées et autonomes.

La continuité du récit peut être caractérisée par des éléments que nous avions considérés comme « sémantique » ou « lexicaux ». La présence d'une liaison a des effets divers : dans le cas d'une clause matrice possédant un mot de liaison, la clause précédente est produite plus rapidement, mais la clause C_n est produite plus lentement, alors que c'est le pattern inverse pour les clauses autonomes. Le fait que le sujet soit maintenu par rapport à la clause précédente entraine généralement une production plus lente, hormis dans le cas où il y a déjà un effet de ralentissement sur la clause précédente, ce qui permet sans doute une anticipation (cas des clauses matrices). De manière surprenante, plus un référent est difficile d'accès, plus il est produit rapidement, cet effet n'étant présent que pour la pause initiale de la clause précédant une clause matrice. Enfin, la continuité du récit se manifeste également par les items lexicaux utilisés pour les termes d'une même chaine référentielle. De manière globale, les répétitions exactes de l'item lexical ou une réinsertion d'un item déjà utilisé entrainent une production plus rapide. En revanche, la proportion d'insertion d'un nouvel item lexical affecte la durée de production des clauses matrices, avec une diminution de la durée pour la pause initiale de C_{n-1}, mais une production plus lente pour l'écriture de la clause elle-même.

Peu de variables concernaient les choix syntaxiques. Nous pouvons cependant noter que les effets vont globalement dans le sens attendu : la présence d'un verbe modal rend la production plus lente, tout comme le fait que la profondeur syntaxique maximum de la clause soit plus élevée (hormis pour la pause initiale de la clause matrice).

Enfin, la dernière série de variables rendait compte d'aspects lexicaux : nombre de séquences, fréquence, nature grammaticale. Les résultats liés au nombre de séquences dépendent du type de clause. En effet, alors que les variables chronométriques sont plus courtes dans le cas d'une clause matrice longue, elles sont plus longues dans le cas d'une longue subordonnée. Les effets de la nature grammaticale sont visibles d'une part à travers la variable « densité lexicale » (pourcentage de mots lexicaux) et d'autre part par rapport à la nature grammaticale du sujet. Les effets de cette dernière ne concordent pas tous, et dépendent du type de clause. Ainsi, le fait que le sujet soit une ellipse ne conduit pas à une temporalité significativement différente pour les subordonnées ; il entraine en revanche une durée plus longue que les noms propres et pronom pour les clauses autonomes, mais une durée plus

courte pour les clauses matrices. En outre, la présence d'un pronom induit une longueur plus longue que s'il s'agit d'un syntagme nominal pour la production de la clause précédant une subordonnée, mais la production est plus courte pour la subordonnée elle-même.

2.4. Discussion générale

2.4.1. Temporalité de la planification des éléments linguistiques

Nos résultats ont d'abord porté sur la présence d'effets significatifs, afin de déterminer quand prenaient place la micro-planification (sémantique) et la formulation (syntaxique et lexicale). Concernant les phrases, le gain explicatif apporté par l'introduction des variables linguistiques n'a pas été très élevé, notamment pour la dernière clause de la phrase précédente. La variation de durée de la pause initiale de la phrase, qui est en moyenne très longue, est finalement peu expliquée par les variables linguistiques de l'ensemble de la phrase. Plusieurs hypothèses peuvent être émises concernant ce résultat : il est tout d'abord possible que l'ensemble de la phrase ne soit pas planifiée dès le début de cette phrase ; dans ce cas, le fait que la plupart des variables linguistiques portent sur l'ensemble de la phrase fournit un résultat moins pertinent que si nous n'avions pris en considération par exemple que les VI liées à la première clause de la phrase. Ainsi, une partie de la micro-planification et de la formulation pourrait être incrémentale, et mieux observée clause par clause. Cependant, si la planification ne porte pas sur un empan si large, la durée de la pause initiale ne pourrait pas être expliquée par ce besoin de planification. Elle pourrait alors être davantage due à des traitements portant sur l'unité précédente comme la relecture (Fortier & Préfontaine, 1994) ou l'intégration en MLT (Daiute, 1984) : ces variations de durée pourraient alors être davantage expliquée par des VI liées à la phrase qui vient d'être écrite. Nous retrouverions ainsi des effets que l'on pourrait qualifier de « retardés », comme ce qui avait été observé par Maggio et al. (2012) dans le cadre de leur étude portant sur le niveau du mot. Cependant, même si l'ampleur des effets est assez faible, les hypothèses concernant le type de VI affectant les différentes VD sont plutôt confirmées : ainsi, la clause de la phrase précédente n'est affectée que par des VI d'ordre sémantique, tout comme la pause initiale. Les VI les plus anticipées portent principalement sur le lien entre les phrases, et sur le prédicat, celui-ci semblant être un élément planifié assez en avance, avant le début de l'écriture de la phrase.

Un pattern ressemblant peut être relevé concernant les clauses autonomes, c'est-à-dire les cas où la clause n'est pas subordonnée ni suivie d'une subordonnée (donc unique clause de l'unité A). Nous observons dans ce cas que les VI expliquent assez peu les variations chronométriques. Les quelques VI ayant un effet significatif sur les VD de la clause précédente concernent également le prédicat, la liaison et en outre la nature grammaticale du sujet : ce sont donc les éléments qui sont planifiés le plus tôt. Il est cependant étonnant que la variation chronométrique de la clause elle-même soit si peu expliquée par les variables linguistiques (entre 13 et 22% de gain explicatif pour les VD internes à la clause, 27% pour

la pause initiale). Il est possible que l'absence de forte dépendance syntaxique avec les clauses entourant la clause autonome facilite la production de celle-ci, permettant d'autres traitements non directement liés à la planification de C_n.

Notre hypothèse issue d'études précédentes prévoyait que les clauses davantage dépendantes syntaxiquement seraient anticipées, c'est-à-dire planifiées dès l'écriture de la clause dont elles dépendent (Immonen & Mäkisalo, 2010; van Hell et al., 2008). Cette hypothèse était renforcée par les résultats de nos données chronométriques, selon lesquels les clauses subordonnées étaient produites plus rapidement, et les clauses matrices étaient précédées d'une pause plus longue que les non matrices. Cela suggérait que la clause subordonnée était anticipée, peut-être dès la pause initiale de la clause précédente. Les résultats présents indiquent que, en effet, les clauses subordonnées sont relativement anticipées, les variables linguistiques expliquant une part assez importante de la variation des variables chronométriques de C_{n-1}, notamment de l'exécution motrice. En outre, contrairement aux autres groupes de clause, les variables syntaxiques (profondeur syntaxique maximum, présence d'un verbe modal) affectent non pas la clause en cours d'écriture, mais la clause précédente. En revanche, un pattern surprenant apparait concernant les clauses matrices. En effet, étant les premières clauses de l'unité A, il n'était pas attendu qu'elles soient particulièrement anticipées, tout comme les clauses autonomes. Or, le gain explicatif dû aux VI de la clause matrice est relativement important pour les variations des VD de la clause précédant la clause matrice ; ce qui n'est pas le cas des clauses autonomes. On pourrait alors envisager que, dès l'écriture de la clause précédente, le scripteur sait qu'il va devoir produire une clause suivie d'une subordonnée. Afin de libérer des ressources pour l'anticipation de la clause subordonnée, il commence alors à planifier sa clause matrice pendant la production de la clause précédente. Cela pourrait peut-être expliquer pourquoi la variation des données chronométriques des clauses autonomes est relativement peu expliquée par les variables linguistiques de celles-ci : le fait qu'elles n'aient pas de lien de dépendance trop fort libère des ressources cognitives, qui peuvent être utilisées pour d'autres traitements, par exemple l'anticipation d'une clause matrice qui suivrait.

2.4.2. Cout de la planification des éléments linguistiques

La difficulté de notre corpus est qu'il n'était pas possible de contrôler le contenu linguistique de ce qui était produit : longueur, fréquence, prédictibilité, etc. Contrairement à des protocoles plus expérimentaux où un facteur peut être précisément étudié, toutes choses étant égales par ailleurs (e.g. Chanquoy et al., 1990), nous avons préféré privilégié des tests statistiques permettant de prendre en compte cette variabilité de contenu, tout en identifiant les variables ayant un effet significatif, nous inspirant ainsi du travail de Maggio et al. (2012) sur le mot produit dans un texte. Afin de faire ressortir au mieux les variables pertinentes, beaucoup de variables ont été codées. Parmi ces variables, certaines n'ont eu aucun effet significatif sur la variation : le pourcentage de syntagmes obligatoires d'un point de vue communicatif et le type de complément de lieu (externe / corporel). Pour des analyses futures,

ces informations ne seront donc pas nécessairement pertinentes, même si leur présence permettait peut-être de contrôler certains effets. Sinon, les autres variables codées avaient des effets significatifs sur une ou plusieurs VD.

Certaines variables ayant un effet significatif semblent être pertinentes à conserver, puisqu'elles expliquent une part de la variance, mais leurs effets diffèrent suivant les types de clauses : c'est par exemple le cas de certaines données liées au prédicat (temporalité, sémantique) et au sujet (nature grammaticale).

En revanche, certains effets restent davantage constants. Tout d'abord, il semble que la saillance discursive conduise à une production plus lente. Ainsi, le fait que le prédicat soit de 1er plan entraine des durées plus longues pour tous les types de clauses et au niveau de la phrase (hormis dans le cas de la clause subordonnée, mais pour laquelle un effet de ralentissement a été observé sur la clause précédente, sans doute dû à une anticipation). De même, les variables chronométriques sont généralement plus longues lorsque le sujet est « Dimitri », c'est-à-dire le personnage principal du récit. Cela pourrait être dû au souci du scripteur de construire un récit cohérent, où chaque étape importante est clairement mentionnée ; on peut rapprocher ces résultats de ceux concernant le lien entre durée de pauses longues et structure hiérarchique du texte (Schilperoord & Sanders, 1999). Ici, ce ne sont pas tant les pauses initiales qui sont affectées que l'ensemble de la production de la clause, mais il serait ainsi possible qu'un ralentissement soit observé au moment de l'insertion de nouvelles macropropositions (au sens de Adam, 2011).

Au niveau de la continuité discursive, des résultats intéressants, et pas forcément attendus, sont apparus. En effet, nous attendions plutôt que le fait qu'une unité soit fortement liée à la précédente accélère sa production, puisque les informations seraient davantage accessibles : en lecture, un indice d'ASL élevé (donc un fort lien entre les unités) conduisait à une lecture du mot plus rapide (Pynte et al., 2008b). En revanche, l'effet inverse est observé en écriture : pour les phrases, un indice d'ASL élevé conduite à une production plus lente de la dernière clause de la phrase précédente. Un maintien du sujet conduit également à un allongement du temps de la production, sauf si un allongement est déjà observé lors de la clause précédente (cas de la clause matrice). Les phrases sont en outre plus longues à produire lorsqu'il y a un adverbe de liaison ; la présence d'un adverbe de liaison conduit à des effets plus contrastés au niveau des clauses (pour les matrices une diminution du temps de C_{n-1} puis un allongement de C_n, et l'inverse pour les clauses autonomes). Ce cout lié au lien entre deux unités peut être dû aux contraintes de cohérence, qui font que le scripteur doit être particulièrement attentif au choix des mots pour maintenir une cohésion dans son texte et une progression pertinente dans l'histoire.

Des effets de fréquence sont également relevés dans le texte : par exemple, un prédicat plus fréquent conduira à une production plus rapide, que ce soit au niveau des phrases ou des clauses autonomes. Cette fréquence des prédicats, calculée sur le corpus, peut s'apparenter à une caractéristique de prédictibilité : plus les prédicats sont fréquents, plus il est attendu que les scripteurs les utilisent, notamment

conformément au script de l'évasion. On retrouve alors un effet de plus grande rapidité, comme ce qu'avaient noté Chanquoy et al. (1990) concernant l'effet de prévisibilité de la fin du texte à écrire. L'effet de fréquence est moins constant sur nos données. En effet, il apparait que le fait qu'un lemme ou qu'un lexème soit fréquent diminue généralement la durée des données chronométriques. Cependant, dans des cas où un tel effet est observé pour la fréquence basée sur Manulex, l'effet est opposé pour la fréquence basée sur le corpus : il semble ainsi que, une fois la fréquence sur la base de données prise en compte, le fait que les items soient fréquents dans le corpus rende la production de la clause plus lente.

Concernant la longueur, un pattern assez complexe émerge : nous avons noté que les variables chronométriques sont plus courtes dans le cas d'une clause matrice longue, mais plus longues dans le cas d'une longue subordonnée. Cela pourrait être dû à des contraintes de mémoire : le scripteur a intérêt à écrire rapidement sa clause matrice, surtout si elle est longue, afin de garder en mémoire cette clause et la suivante qui a commencé à être planifiée ; en revanche, dans le cas d'une subordonnée cette contrainte est moins importante et la production est donc plus longue si la clause est plus longue. Cette contrainte de maintien en mémoire pourrait également expliquer pourquoi, lorsqu'il y a beaucoup de référents dans la clause subordonnée, l'écriture de la clause précédente est plus rapide.

Concernant la syntaxe, on s'attendait à ce qu'une plus grande complexité entraine un cout plus élevé : c'est globalement le cas, puisque la présence d'un verbe modal allonge les données chronométriques, tout comme le fait que la clause soit très profonde (hormis pour la clause initiale de la clause matrice).

Enfin, certaines caractéristiques spécifiques à notre type de récit présentent des résultats réguliers, même si nous n'avions pas beaucoup d'hypothèses sur ces variables. Ainsi, il ressort que la présence d'un complément de lieu allonge le temps de production de la clause. La mention des lieux est en effet importante dans nos récits d'évasion. Cependant, beaucoup de choix s'offrent au scripteur concernant ces lieux : par exemple, lorsqu'il s'agit d'un lieu « corporel » (e.g. la tête du gardien, ses jambes), le scripteur doit déterminer quel lieu est visé (e.g. « tête », « nuque ») et peut avoir différentes manières de désigner la personne (e.g. « du gardien », « sa ») ; par ailleurs, le choix d'un lieu « externe » (e.g. « dans le couloir », « dans les escaliers ») détermine fortement les étapes de la narration, et doit donc être fait avec soin. À l'inverse, la présence de compléments de temps « balises » accélère le récit. Ces compléments rythment le texte, mais sont assez faciles à choisir : « puis », « ensuite », « le lendemain » ; leur production est donc assez fluide, permettant à la clause d'être produite plus rapidement. Il est à noter que ce résultat est contraire à celui de Bertrand et Dumoulin (2013), qui notaient que dans les récits d'expériences du projet Spencer, la pause initiale avait tendance à être plus longue lorsque l'unique cas nouveau introduit était une information temporelle que s'il s'agissait d'un complément de lieu : cela pouvait alors s'expliquer par le fait que, pour ces récits, les lieux étaient habituels pour les scripteurs (e.g. maison, collège), alors que les informations temporelles étaient plus difficiles d'accès. Enfin, nous pouvons relever que, dans nos récits d'évasion, le fait d'insérer davantage de référents « personnages »

rend la production plus rapide, sans doute parce que ce sont des participants faciles à identifier et dont les rôles sont clairement définis au sein de la clause. L’insertion de ces variables linguistiques spécifiques à notre récit est donc pertinente pour expliquer la variance des données chronométriques, mais leur réplicabilité n’est pas forcément attendue dans le cadre de la production d’autres types de textes.

3. Conclusion

Suite aux études suggérant un effet d'anticipation de certaines clauses en fonction de leur degré de dépendance syntaxique, nous souhaitions vérifier dans quelle mesure certains éléments linguistiques des phrases et des clauses étaient planifiés dès l'écriture de la clause précédente. En outre, peu d'études examinaient quelles caractéristiques linguistiques provoquaient un cout plus important, ou au contraire moins important, de la production. C'est pourquoi nous avons codé de très nombreuses variables linguistiques, qui ont été insérées dans des modèles afin d'expliquer une part de la variance des données chronométriques des clauses en cours d'écriture et des clauses précédentes. Outre les variables linguistiques insérées dans le modèle, la variance des données chronométriques pouvait également être expliquée par les caractéristiques des participants, ce qui a conduit à faire des modèles mixtes dans lesquels la variable « participant » constituait un facteur aléatoire. L'ICC permettait de rendre compte la part de variance restant pouvant être expliquée par les participants. De manière générale, les résultats ont montré que cet ICC était (très) élevé pour le temps d'exécution ; cela rejoint notamment la description chronométrique qui avait fait émerger deux profils de scripteurs concernant le pourcentage de temps consacré à l'exécution (10% ou 20%). En revanche, la variance des pauses initiales des clauses est généralement moins expliquée par les variations interindividuelles.

Les résultats issus de l'analyse préliminaire présentée dans le chapitre 6 suggéraient de faire des modèles pour différents groupes : phrase, clause matrice, clause subordonnée, clause autonome. Cette distinction parait pertinente, puisque des patterns différents émergent. Le fait d'étudier des clauses dans le contexte d'un texte a permis de faire apparaitre des effets qui ne se seraient certainement pas manifestés lors de la production de phrases : par exemple, l'anticipation de la clause matrice lors de l'écriture de la clause précédente est certainement due à une gestion temporelle des processus au niveau textuel, permettant d'anticiper un cout plus important à venir. En outre, des résultats spécifiquement liés à la continuité discursive ont ainsi pu émerger, comme le fait que le rythme de production soit globalement ralenti lorsqu'une unité est davantage liée à la précédente au niveau sémantique.

Certains résultats présentés ici sont complexes à analyser. En effet, la volonté de prendre en compte beaucoup de variables linguistiques a permis de faire émerger des effets significatifs, mais ces derniers ne présentent pas toujours un pattern régulier (e.g. temporalité de l'écriture, nature grammaticale du sujet, type de reprise lexicale). Parfois, certains résultats complexes ont pu être expliqués en prenant en compte le fait qu'un effet soit déjà apparu sur la clause précédente, provoquant par exemple un effet facilitateur sur la clause en cours (e.g. la continuité du sujet pour les clauses matrices), ou en prenant en compte la présence d'un effet significatif simultané d'une VI assez proche (e.g. fréquence basée sur le corpus et sur Manulex). Il sera peut-être nécessaire de recourir à des protocoles expérimentaux plus contraints ou à des études sur un corpus plus large pour éclaircir certains résultats.

Chapitre 8 – Étude des révisions

La mise en relation du temps d'écriture des phrases et des clauses avec le contenu linguistique de celles-ci a permis de mieux saisir quand les éléments étaient planifiés et le cout cognitif de leur planification ou production, mais elle n'a pu s'effectuer que sur des séquences de texte écrites de manière linéaire, donc sans séquences de révision. Or, les révisions peuvent également donner des informations sur la micro-planification et la formulation. Dans certains cas, si le texte comporte une très grande quantité de révisions, il est même biaisé de généraliser la manière dont il a été construit en n'étudiant que les séquences sans révision. Pour approfondir davantage les processus rédactionnels grâce à l'écriture enregistrée, ce dernier chapitre explorera les révisions des textes.

Dans un premier temps, les révisions présentes dans le corpus « Dimitri », utilisé dans les chapitres précédents, permettront d'affiner certaines hypothèses. En effet, les révisions n'étaient pas très nombreuses dans ce corpus (entre 1 et 15 par texte), ce qui permettait d'avoir suffisamment de données sur des séquences « linéaires ». Nous nous servirons donc de ces révisions comme indices supplémentaires quant à l'empan de planification et à la gestion des processus rédactionnels au sein de ces unités de production. Dans un deuxième temps, l'étude d'un texte comportant un très grand nombre de séquences de révision montrera dans quelle mesure celles-ci fournissent des indices très précieux pour mieux comprendre l'élaboration de ce récit.

1. Révisions du corpus

1.1. Codage des révisions

Les révisions des textes ont été relevées et codées. Comme évoqué dans le cadre théorique, de nombreuses définitions de la « révision » peuvent être apportées. Les définitions retenues et le codage apporté peut dépendre d'une part de la question de recherche, et d'autre part du type de données recueillies. Dans le cas de notre corpus, il s'agissait de produits manuscrits, et les données été enregistrées grâce au logiciel Eye & Pen : il est ainsi possible de savoir si le stylo a été déplacé sur des séquences antérieures ; si c'est le cas, nous pouvons connaitre la chronologie relative (ordre des révisions) et la chronologie absolue (moment de la révision dans le texte) ainsi que les durées associées à ces séquences de révision.

Nous avons appelé « révision » toute opération ayant laissé une trace de retour sur le texte déjà écrit, c'est-à-dire sur une séquence autre que celle en cours d'écriture. Nous parlons de « trace de retour » et non de « modification » car nous avons relevé de très nombreux cas de « révision » où le texte n'est pas modifié. Ce type d'opération, rarement évoqué dans la littérature, a été nommé pour notre étude « reprise esthétique », car il semblait que le bu premier de ces retours sur le texte était d'améliorer la lisibilité de celui-ci. L'absence de typologie concernant cette opération peut être expliquée par le matériau utilisé : ce type de retour sur le texte est invisible sur les manuscrits, et ne peut pas exister dans le cas du traitement de texte.

Plusieurs informations permettent de caractériser les révisions (Rey-Debove, 1981). Le but de cette étude étant de mieux comprendre la gestion temporelle des processus rédactionnels grâce aux révisions, nous avons codé d'une part le type d'opération de révision, et d'autre part des informations de localisation concernant la révision (distance entre l'origine de la révision et le site de la révision[69] ; position de la révision dans le texte).

1.1.1. Type d'opération de révision

Plusieurs opérations de révision peuvent être identifiées. Nous nous sommes inspirés de la classification et des remarques proposées par Doquet et Leblay (2014). Une réflexion quant à la classification de ces opérations sera proposée lors de leur analyse. Doquet et Leblay (2014) identifient quatre opérations : suppression, ajout, remplacement et déplacement, cette dernière opération n'ayant pas été relevée dans notre corpus. Nous avons en outre ajouté un autre type d'opération : la « reprise

[69] Le « site » correspond à l'élément affecté (ex : le segment qui est raturé) ; l' « origine » correspond au lieu où le scripteur s'arrête pour effectuer sa révision (appelé dans certaines études « point d'inscription »).

esthétique ». Ces quatre types d'opération pouvaient porter sur un signe de ponctuation (dans le codage « pct ») ou un mot[70].

- Suppression (« SUP ») : Il s'agit de la suppression, généralement visible sur le manuscrit, d'un segment préalablement écrit (138).

  (138) fut effrayé du regard si noir et intense que j **SUP « j »** lui jetait le détenu[71] [S26]

- Ajout (« AJ ») : Il s'agit d'un retour sur le texte durant lequel un segment est inséré alors qu'aucun élément syntagmatiquement équivalent (dans l'exemple (139), un signe de ponctuation) n'était présent à ce site (ici, entre « agir » et « c' »).[72]

  (139) *S'il voulait agir __ c'* **AJ « , »** était maintenant ou jamais [S07]

- Remplacement (« REMP ») : Les cas de « remplacement » sont assez rares, et pourraient être confondus avec une suppression suivie d'un ajout. Nous avons considéré qu'une révision correspondait à un « remplacement » lorsqu'un segment est supprimé puis qu'il y a immédiatement un ajout d'un segment syntagmatiquement équivalent.

  (140) qu'il avait eue avec Dimitri, deux jours auparavant, alors que **REMP « , > . »**[73]

- Reprise esthétique (« ESTH ») : Les « reprises esthétiques » désignent des retours sur le texte laissant une trace visible lors du déroulement de la séance avec le logiciel Eye & Pen mais n'apportant aucune modification. Dans ces cas, le scripteur ne fait généralement que repasser un segment (par exemple, un point sur un « i », ou une lettre).

En outre, nous avons isolé les modifications orthographiques (« ORTHO ») : Il s'agit d'opérations visant à corriger l'orthographe, lexicale ou grammaticale, d'un mot. Certes, celles-ci pourraient être identifiées grâce aux opérations ci-dessus, puisque pour corriger un mot on peut supprimer une ou plusieurs lettres (ex : plans > plan), en ajouter (ex : leur > leurs) ou effectuer un remplacement (ex : voie > voix). Cependant, la question de l'ajout et de la suppression est difficile d'un point de vue morphologique. Par exemple, dans le cas d'une correction de genre d'un adjectif, on aurait un remplacement dans le cas de « veuf > veuve » ou « veuve > veuf », mais un ajout pour « petit > petite »

[70] Théoriquement, il pourrait également s'agir d'un groupe de mots. Cependant ce cas n'est relevé qu'une fois dans notre corpus, et il s'agit en outre d'une locution (« alors que »), nous l'avons donc comptabilisé comme « mot ».

[71] Dans chaque exemple, la révision est indiquée en gras, et le site est souligné.

[72] Contrairement à Doquet (2003) nous ne comptons pas l'inscription de segments à la suite de ce qui est écrit comme une opération d' « ajouts » à étudier, puisque nous ne nous intéressons qu'aux révisions.

[73] le signe « > » indique le changement effectué (= « devient »)

et une suppression pour « petite > petit ». De plus, il est difficile de dire dans le cas d'une modification orthographique si l'élément modifié est précisément le mot ou une unité infra-lexicale (morphème, lettre, accent). Il nous est donc paru plus pertinent, lors du codage, de considérer une seule opération.

1.1.2. Localisation de la révision

1.1.2.1. Distance entre l'origine et le site de la révision

Nous avons noté la distance entre l'origine de la révision (le lieu où le scripteur s'est arrêté et est revenu en arrière pour apporter une modification) et le site de la révision (le segment modifié).

La révision pouvait portait sur le mot en cours (« n ») (141), ce qui était uniquement identifiable dans le cas d'une suppression de mot. En effet, si le scripteur apportait une reprise d'ordre esthétique ou modifiait une lettre du mot au cours de l'écriture de celui-ci, l'ensemble des opérations étaient comptabilisées dans une seule séquence « d'écriture du mot » : ces opérations n'étaient donc pas considérées comme des « révisions ». Sinon, la révision pouvait porter sur le mot précédent (« n-1 ») (142), sur un mot situé dans la même clause (« c », pour « courte ») (143) ou sur un mot situé dans une autre clause, mais effectuée au cours de l'écriture du texte (« m » pour « moyenne »). Enfin, il pouvait arriver que des révisions soient effectuées à la fin de l'écriture du texte (« l » pour « longue »).

(141) mais Amin était prêt à tout pour sauver son frère Tho **SUP «Tho »** Dimitri de la peine de mort. [S01]

(142) Dimitri était dans un couloir __ il **AJ « , »** faisait très sombre [S17]

(143) Alors Dimitri poussa de toute ses forces **ORTHO « toute > toutes »** [S24]

1.1.2.2. Position de la révision dans le texte

Pour chaque révision, nous avons localisé la position de l'origine et du site de la révision. Pour cela, nous avons indiqué :

- la partie du texte (telle que segmentée en cinq parties pour les études chronométriques) ;
- la position de la séquence par rapport au début de la phrase, à la fin de la phrase, au début de la clause et à la fin de la clause. La première séquence de la phrase est le premier mot ; la première séquence par rapport à la fin de la phrase est le signe de ponctuation qui clôture cette phrase.

Par exemple, la phrase (144) contient une modification orthographique : « toute > toutes ». Le site de cette révision (« toute ») est la 5ème séquence par rapport au début de la phrase et la 9ème par rapport à la fin de la phrase ; c'est la 5ème séquence par rapport au début de la clause et la 3ème par rapport à la fin de la clause. L'origine de la révision (fin du mot « forces ») est la 7ème séquence par rapport au début de la phrase et la 6ème par rapport à la fin de la phrase ; c'est la 7ème séquence par rapport au début de la

clause et la 1ème par rapport à la fin de la clause. Ainsi, le site de cette révision est situé au milieu de la phrase et de la clause, mais la révision est initiée en fin de clause.

(144) Alors Dimitri poussa de <u>toute</u> ses forces **ORTHO « toute>toutes »** et la grille *s'arracha. [S24]*

1.2. Hypothèses

Cette étude des révisions n'a pas pour objectif de décrire dans le détail les révisions effectuées, mais de voir dans quelle mesure le lien entre type d'opération, localisation de la révision et durée des révisions permet de mieux comprendre la gestion temporelle des processus rédactionnels. Cela permettra notamment d'une part de corroborer ce qu'a montré le travail présenté précédemment quant aux unités de production ; d'autre part de comprendre où se situent certaines difficultés liées à la micro-planification et à la formulation, et quand elles sont résolues.

Tout d'abord, il est attendu, conformément à ce qui avait été observé expérimentalement par Quinlan, Loncke, Leijten et van Waes (2012), que les révisions soient initiées à la fin d'une unité de production. En outre, la longue pause entre phrases, peu expliquée par la planification de la phrase à venir, serait davantage due à un temps de relecture

H1 : Les révisions seraient majoritairement initiées en fin d'unités linguistiques.

Par ailleurs, même si certains éléments de la planification (notamment le prédicat) sont planifiés avant l'écriture de l'unité, une part de la planification parait être incrémentale. Des hésitations pourraient donc avoir lieu pendant l'écriture de l'unité, et seraient visibles à travers des modifications du choix initial.

H2 : Des reformulations liées au lexique pourraient avoir lieu pendant l'écriture de la phrase, et donneraient lieu à des modifications immédiates.

1.3. Description générale

En tout, le corpus « Dimitri » compte 142 révisions, soit en moyenne un peu moins de 5 par texte. Cela masque quelques différences suivant les scripteurs, comme illustré dans la Figure 55 : 19 scripteurs font moins de 5 révisions, 5 scripteurs en font entre 5 et 9 et 6 scripteurs en font entre 10 et 15.

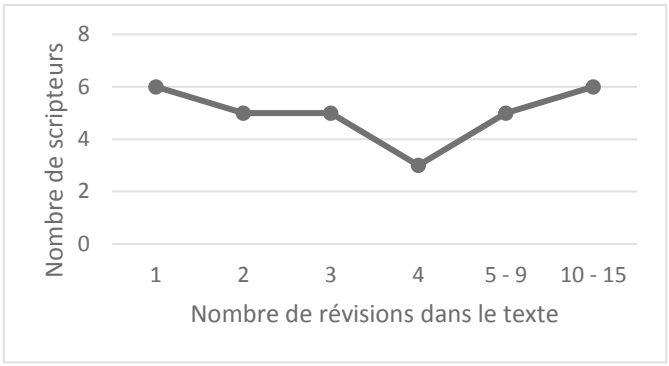


Figure 55. Distribution des scripteurs en fonction du nombre de révisions dans le texte

Suivant les cinq opérations décrites dans le codage, la Figure 56 montre que l'opération la plus fréquente est la reprise esthétique (35%), alors que la plus rare est le remplacement (4%) ; il y a en outre légèrement plus d'opérations de suppression (20%) que d'ajout (16%). On peut également regarder plus précisément le type d'élément modifié, comme illustré par la Figure 57 : ainsi, il est possible de voir que l'attention du scripteur se porte majoritairement sur les mots (57%), et moins sur les signes de ponctuation (18%).

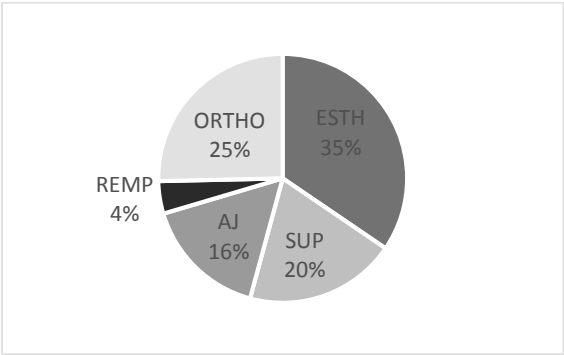


Figure 56. Répartition des révisions en fonction des types d'opérations (en %)

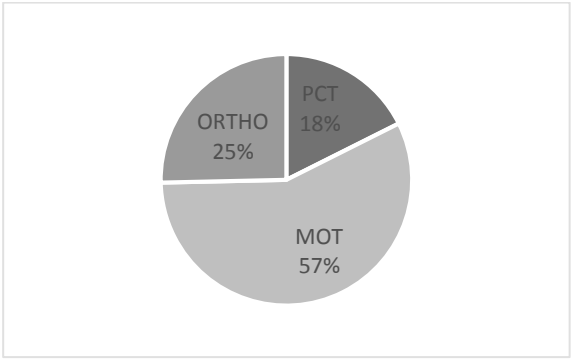


Figure 57. Répartition des révisions en fonction du type d'éléments modifiés (en %)

La Figure 58 montre que quasiment toutes les opérations de « reprise esthétique » portent sur le mot (47 sur 49), tout comme les opérations de suppression (27 sur 28), alors que les scripteurs ajoutent davantage de signes de ponctuation que de mots (19 sur 23) ; les cas de remplacement, plus rares, portent autant sur les mots que sur la ponctuation.

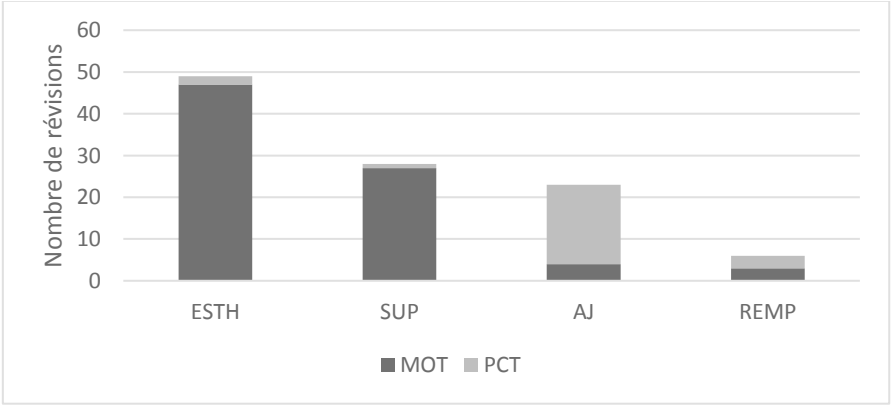


Figure 58. Type d'opération de révision en fonction de l'élément révisé

1.4. Localisation des révisions

Afin de conserver la précision de nos données tout en évitant de généraliser à partir de données trop peu nombreuses, nous avons réorganisé les catégories pour la présentation de ces résultats. Ainsi, nous avons souhaité conserver l'information quant à l'opération « reprise esthétique », qui est une opération très fréquente et distincte des autres, puisque le scripteur ne modifie pas le texte déjà écrit. Par ailleurs, parmi les opérations, il y a très peu de cas de remplacement, et les opérations d'ajout et de suppression sont associées en grande partie respectivement aux éléments révisés « ponctuation » et « mot ». Les quatre catégories retenues sont donc :

- « ORTHO »
- « ESTHE » (qui porte presque toujours sur un mot)
- « MOT » (hormis les cas de reprise esthétique, donc presque toujours des suppressions)
- « PCT » (hormis les cas de reprise esthétique, donc presque toujours des ajouts)

1.4.1. Distance entre l'origine et le site de la révision

Le Tableau 63 indique le nombre de révisions en fonction de sa catégorie et de la distance entre l'origine et le site de la révision. La Figure 59 et la Figure 60 illustrent cette distribution en termes de pourcentages.

	ESTH	MOT	ORTHO	PCT	Total
Longue (l)	0	1	3	2	4%
Moyenne (m)	5	2	4	1	8%
Courte (c)	7	2	11	2	16%
Mot précédent (n-1)	37	5	18	18	55%
Mot en cours (n)	0	24	0	0	17%

Tableau 63. Distance entre l'origine et le site de la révision en fonction des catégories de révision

Peu de révisions (12%) portent sur des séquences situées avant la clause en cours d'écriture ; en majorité, les révisions portent sur le mot précédent (55%). Comme attendu suite à notre manière de coder les révisions, les seuls cas de révisions sur le mot en cours (« n ») consistent en des suppressions de mot, ces suppressions immédiates constituant 70 % des révisions portant sur les mots. Pour les autres types de révision (ESTH, ORTHO, PCT), entre 50 et 80% des opérations portent sur le mot précédent. Les modifications orthographiques se font aussi assez souvent (30%) au sein de la clause ; il s'agit en outre de l'opération la plus fréquente dans le cas des révisions « longues », c'est-à-dire effectuées à la fin du texte (3 cas sur 6). Les révisions effectuées sur la clause en cours d'écriture ou plus loin dans le texte déjà écrit consistent essentiellement en des modifications orthographiques ou des reprises esthétiques.

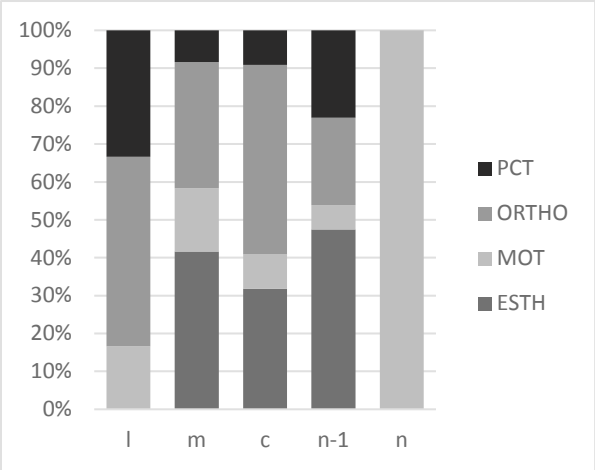


Figure 59. Distribution des catégories de révision en fonction de la distance entre l'origine et le site (en %)

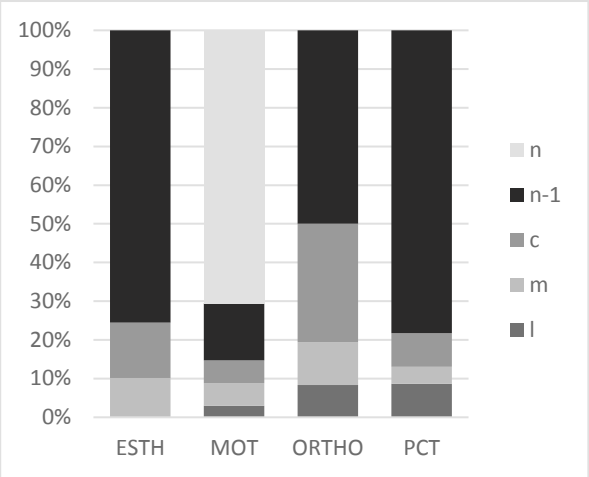


Figure 60. Distribution la distance entre l'origine et le site en fonction des catégories de révision (en %)

1.4.2. Localisation dans le texte

1.4.2.1. Localisation par rapport aux parties du texte

Deux éléments doivent être distingués lorsque l'on s'intéresse à la localisation d'une révision : où se situe l'élément modifié (le « site ») et à quel endroit la révision est initiée (l'« origine »). Le texte a été divisé en 5 parties de durée équivalente. Le Tableau 64 indique le nombre d'origines et sites de révision dans chaque partie, ainsi que leur distribution en fonction des quatre catégories de révision ; ces distributions sont illustrées par la Figure 61 et la Figure 62. La partie la plus touchée par les révisions est la partie 3, que ce soit pour les sites ou les origines des révisions. De manière globale, toutes les parties des textes sont touchées par des révisions, même si les sites des révisions sont légèrement moins nombreux aux parties extrêmes (1 et 5). Du fait que certaines révisions soient différées, il y a en général moins d'origines de révisions que de sites, hormis dans la partie 5, où il y a davantage de révisions initiées : c'est notamment dans cette partie que sont effectuées les révisions « longues ».

La répartition des catégories de révision au cours du texte est quant à elle assez stable. Si l'on se réfère aux pourcentages illustrés par la Figure 62 en comparant les parties entre elles, on peut noter une proportion un peu plus élevée de reprises esthétiques dans la partie 2, un peu plus de révisions portant sur le mot dans la partie 3, davantage de révisions sur la ponctuation dans la partie 4 et une part plus importante de modifications orthographiques dans la partie 5.

	site	origine	ESTH	ORTHO	MOT	PCT
partie 1	24	21	7	7	7	3
partie 2	28	26	12	6	5	5
partie 3	37	37	11	9	11	6
partie 4	29	25	10	6	6	7
partie 5	24	33	9	8	5	2

Tableau 64. Localisation des révisions en fonction des parties du texte

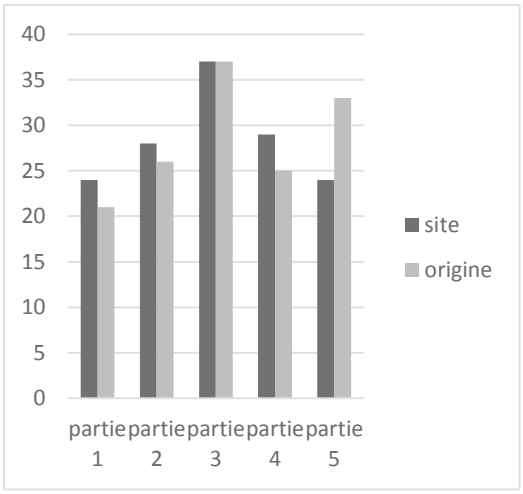


Figure 61. Localisation des sites et origines des révisions en fonction des parties du texte

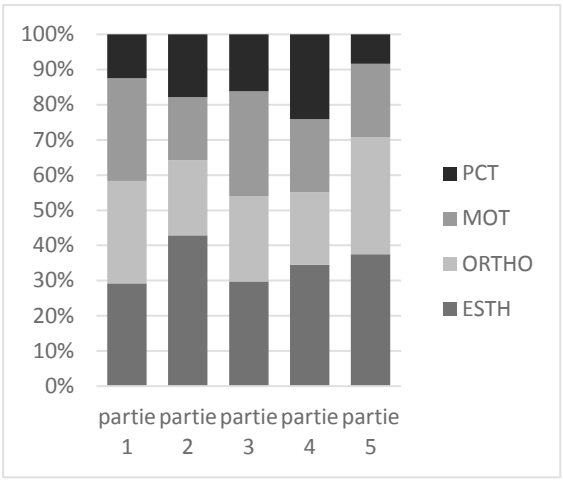


Figure 62. Localisation des catégories de révisions en fonction des parties du texte

1.4.2.2. Localisation par rapport aux phrases et aux clauses

Nous avons regardé dans quelle mesure il était possible de trouver des schémas réguliers concernant la temporalité des révisions : à quel endroit se situent les séquences révisées, et surtout à quel moment les révisions sont initiées. Pour cela, nous avons notamment regardé où se situait l'origine de la révision par rapport à la fin de la phrase ou de la clause. Si elle se situait en « 1ère position par rapport à la fin de la phrase », cela signifiait qu'elle était initiée après le signe de ponctuation, alors que si elle se situait en « 2ème position par rapport à la fin », cela signifiait qu'elle était initiée avant le signe de ponctuation ; dans ces deux cas, la révision avait lieu après l'écriture de tous les mots de l'unité.

Les révisions « ESTH » sont majoritairement initiées en fin d'unité : dans 53% des cas la révision est initiée en fin de phrase, auxquels il faut ajouter 12% de cas où elle est initiée en fin de clause, sans que celle-ci soit la dernière de la phrase. Sur ces 65% d'origine en fin d'unité, la révision est davantage initiée après le signe de ponctuation (80%) qu'entre le dernier mot écrit et le signe de ponctuation. Comme beaucoup de révisions « ESTH » portent sur le mot précédent, il est de manière logique très fréquent que cette révision porte sur le dernier mot de la clause écrite.

Concernant les révisions « PCT », elles sont initiées dans 59% des cas après le premier mot de la clause (dont seulement 9% de cas où c'est également le début de la phrase). Il s'agit généralement d'ajouts, qui sont insérés après le dernier mot de la clause précédente. Il y a 23% de cas supplémentaires dans lesquels la révision portant sur la ponctuation a lieu après avoir écrit le signe de ponctuation clôturant la phrase.

Les révisions « ORTHO » sont quant à elles initiées dans 36% des cas en fin de phrase, que ce soit à part égal après ou avant le signe de ponctuation. Aucun schéma particulier ne ressort concernant les révisions « MOT ».

Les résultats concernant la position de l'origine de la révision en fonction de la distance entre l'origine et la séquence révisée indiquent qu'une majorité des pauses « courtes » (50%) ou « moyennes »

(58%) sont initiées en fin de clause. Cette proportion est plus faible pour les révisions de très courte distance, portant sur le mot (seulement 17%) ou le mot précédent (37%).

1.5. Gestion temporelle des révisions

La présentation générale des données chronométriques avait montré qu'un certain nombre de longues pauses étaient dues à des révisions. Le comportement particulier inhérent à l'inscription de révisions (relecture, déplacement du stylo) nous avait conduits à exclure les séquences comportant des révisions lors de nos analyses portant sur la temporalité. À présent que les révisions ont été plus finement codées et décrites, il serait intéressant de voir dans quelle mesure leurs caractéristiques expliquent la durée des pauses avant l'origine de la révision ou avant la révision elle-même. Afin de prendre en compte la variabilité entre les scripteurs (par exemple, une pause de 1 seconde peut parfois être dans le 3ème quartile, et parfois dans le dernier décile), il a été choisi de décrire les données temporelles non pas en millisecondes mais en termes de quartiles, tels que les seuils avaient été définis lors de la description chronométrique générale.

Le Tableau 65 indique la répartition des pauses dans les quatre quartiles (pour rappel, « Q4 » correspond par exemple aux 25% de pauses les plus longues) en fonction de l'opération de révision, du type de segment modifié (mot ou ponctuation) et de la distance entre l'origine et le site de la révision. Sont distinguées ici les pauses « avant l'origine », c'est-à-dire la durée de la pause initiale du mot après lequel la révision est initiée, et la pause « avant la révision », c'est-à-dire la durée entre la fin du dernier mot écrit et le début de la révision. Sur les 142 révisions relevées dans notre corpus, 46 pauses (32%) avant les révisions appartiennent au décile D10, c'est-à-dire aux 10% de pauses les plus longues de leur texte, leur durée se situant entre 1 seconde et 55 secondes. En revanche, seules 13 pauses avant l'origine des pauses appartiennent au décile D10. De manière générale, il peut être noté que la part de pauses longues (appartenant à Q4) est globalement élevée : c'est le cas de 61% des pauses avant les révisions, et 32% des pauses avant l'origine des révisions

La Figure 63 et la Figure 64 permettent de faire ressortir quelques résultats quant à la gestion temporelle des révisions. Tout d'abord, il apparait que les pauses avant la révision sont plus longues que celles avant l'origine de la révision, à l'exception de deux cas : pour les catégories « AJ » et « PCT ». Or, la plupart des opérations portant sur la ponctuation sont des ajouts, ces deux résultats sont donc liés et suggèrent que l'ajout d'un signe de ponctuation demande généralement plus de temps avant l'écriture du dernier mot qu'avant la révision elle-même. Les modifications orthographiques sont caractérisées par des pauses avant la révision très longue (c'est la catégorie qui a la plus grande proportion de pauses Q4), mais peu de pauses Q4 avant l'origine de la révision. La plus grande proportion de pauses courtes avant l'origine se retrouve dans la catégorie « ESTH ». Par ailleurs, il peut être remarqué que toutes les révisions consistant en un remplacement de mot ou de ponctuation sont précédées d'une pause Q4. En outre, il apparait plus rapide d'initier une suppression qu'un ajout.

Concernant l'impact de la distance sur la durée des pauses, illustré par la Figure 65, on observe un schéma régulier pour les pauses avant les révisions de plus longue distance : il y a plus de pauses Q4 avant une révision « longue », qu'avant une révision « moyenne », qu'avant une révision « courte ». En outre, ce sont les mots à l'origine des révisions « longues » qui sont précédés de la plus grande proportion de pauses longues. La plus grande proportion de pauses courtes se retrouve pour les pauses portant sur le mot précédent.

		Pause avant origine				Pause avant révision			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Esthétique – mot	m		1	1	3			1	4
	c	5	1	1				1	6
	n-1	17	7	6	5	3	5	10	17
Esthétique – ponctuation	n-1				2			1	1
Ajout - mot	l				1				1
	c		1						1
	n-1			1	1	1		1	
Ajout - ponctuation	l		1		1				2
	m			1					1
	c		1					1	
	n-1			4	11	4	5	2	4
Suppression - mot	c				1				1
	n	4	3	6	11	1	2	5	16
	n-1				2			1	1
Suppression - ponctuation	n-1				1	1			
Remplacement - mot	m			1	1				2
	n-1	1							1
Remplacement - ponctuation	c			1					1
	n-1	1			1				2
Modification orthographique	l	1			2				3
	m	1	2		1				4
	c	3	2	5	1			1	10
	n-1	6	6	5	1		2	8	8

Tableau 65. Répartition des pauses avant l'origine et la révision en fonction du type d'opération, de la distance entre l'origine et le site et de la durée de la pause (en quartile)

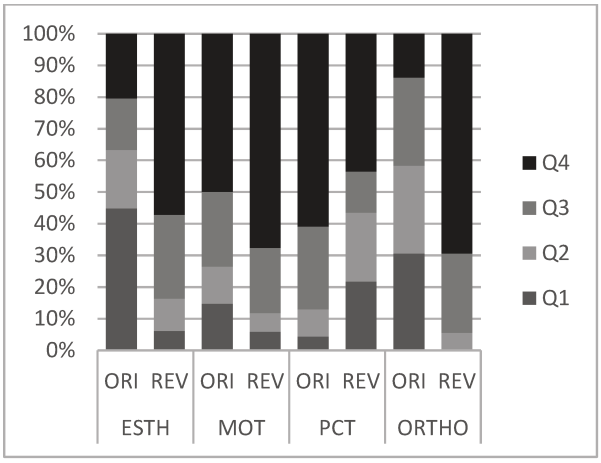


Figure 63. Répartition des pauses en fonction des catégories de révision (durée en quartile)

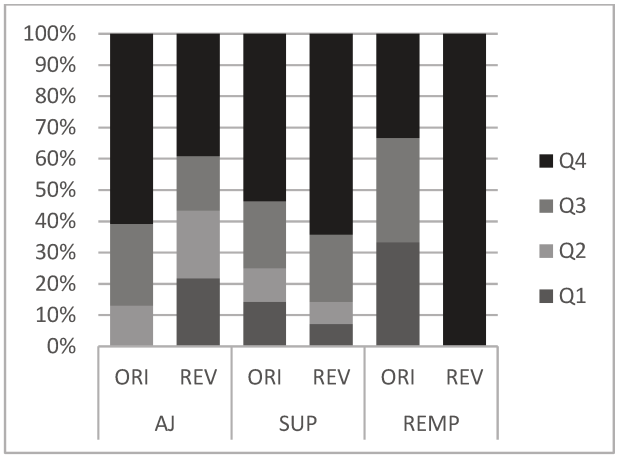


Figure 64. Répartition des pauses en fonction des opérations de révision (durée en quartile)

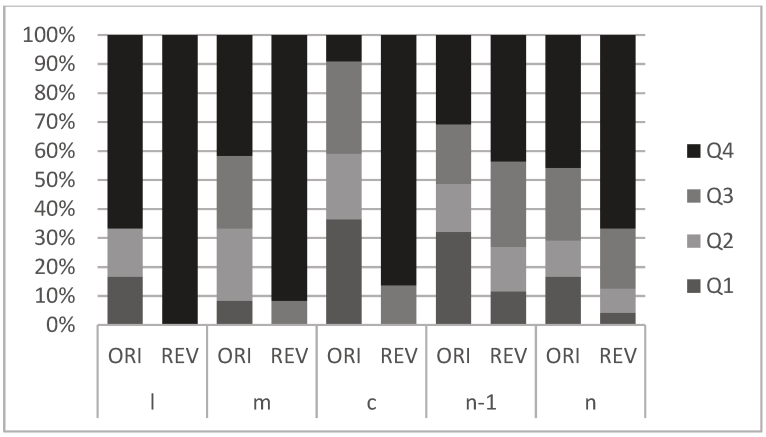


Figure 65. Répartition des pauses en fonction de la distance entre l'origine et le site de la révision (durée en quartile)

1.6. Discussion

1.6.1. La clause, unité de révision ?

Dans le corpus « Dimitri » étudié ici, 88% des révisions portent sur un élément situé dans la même clause. Il y a globalement peu de révisions en fin de texte : seulement 6 dans l'ensemble du corpus, dont 4 effectuées par le même scripteur. Pourtant, 250 erreurs ont été relevées dans le corpus, dont près de 90% n'ont pas été corrigées, chaque texte comportant au moins une erreur. Il n'est donc pas possible de savoir si les 26 scripteurs n'ayant pas fait de révision en fin de texte n'ont pas relu celui-ci à la fin de sa production, mais il est au minimum possible de noter que les erreurs d'orthographes n'ont pas été systématiquement corrigées en fin de production, bien que la consigne stipulât que le texte devait contenir le moins d'erreurs d'orthographe possible, et qu'il était possible de relire le texte une fois la rédaction de celui-ci achevée. Il est donc difficile d'affirmer qu'une relecture systématique soit effectuée en fin de clause, du moins dans le but de contrôler les erreurs d'orthographe. En effet, celles-ci sont corrigées dans 50% des cas dès le mot suivant l'erreur. Dans les cas où elle n'est pas corrigée immédiatement, l'erreur est corrigée 47% du temps à la fin de la clause (avant ou après le signe de

ponctuation) ; sinon, c'est généralement au cours de l'écriture de la clause que l'erreur est repérée et corrigée. Il peut arriver que l'orthographe d'un mot soit modifiée plus loin dans le texte à cause d'un mot proche qui incite à la vérification : par exemple, dans un des textes le scripteur écrit « se rua », puis dans la clause suivante « il porta », et à ce moment s'interrompt pour ajouter un « t » à « se rua » puis inscrit un « t » à la fin de « porta », avant de poursuivre son récit. Dans ce cas le scripteur modifie à tort l'orthographe du premier verbe, mais on peut s'attendre à ce que ce même mécanisme puisse parfois conduire à une correction justifiée. Cet exemple montre en outre qu'il arrive que les scripteurs interrompent le mot en cours d'écriture pour corriger une erreur repérée : c'est le cas dans 25% des modifications orthographiques de notre corpus. Cela rejoint des études utilisant les mouvements oculaires montrant que les scripteurs étaient capables de détecter une erreur en parallèle de l'écriture d'un mot en cours, même si la correction elle-même était parfois différée (Alamargot et al., 2006). Les études portant sur la coordination entre correction orthographique et complétion de la phrase avaient montré que dans 90% des cas les scripteurs préféraient compléter la phrase avant de corriger l'erreur, même s'ils avaient repéré l'erreur avant (van Waes, Leijten & Quinlan, 2010). Il était en outre plus rare de corriger l'erreur avant de compléter la phrase (que ce soit spontanément ou à la demande de l'expérimentateur) si la charge cognitive liée aux mots à retenir ou à la difficulté de la correction était plus élevée. Ce résultat n'est pas conforme à nos observations, puisqu'il est visible que les scripteurs choisissent très largement de corriger l'erreur au cours de la clause, donc sans doute rapidement après l'avoir repérée. Cette tendance à corriger les erreurs immédiatement pendant l'écriture de la clause réfère au processus d'editing, qui intervient de manière fréquente et « automatique » pour corriger rapidement des erreurs repérées (Hayes & Flower, 1980; Kaufer et al., 1986). Le fait que les résultats de l'étude expérimentale diffèrent de ceux relevés dans notre corpus peut être dû aux différences de protocole : dans le cas des expérimentations portant sur la complétion de phrase mentionnées ci-dessus, les scripteurs devaient garder en mémoire les mots à utiliser, qui lui étaient imposés par l'expérimentateur. La préférence pour une correction immédiate dans le cas de notre corpus peut alors avoir deux explications. Une première possibilité est que le scripteur sait parfaitement ce qu'il souhaite écrire pour terminer sa clause ; dans ce cas, il peut se permettre de relire en parallèle ce qu'il vient d'écrire afin de détecter des erreurs, et d'interrompre sa rédaction sans être gêné pour la suite de l'écriture. Une deuxième possibilité est qu'au contraire le scripteur hésite sur la formulation de sa clause ; dans ce cas, il peut choisir de relire ce qu'il vient d'écrire afin d'avoir des idées pour la suite, et s'il détecte une erreur, il choisit de la corriger avant de poursuivre sa rédaction et sa réflexion. L'observation de la durée des pauses peut fournir des indices pour choisir entre ces possibilités. En effet, il a été remarqué que la plupart des pauses avant les révisions étaient longues (Q4). Certes, cette durée comprend le temps pour aller jusqu'au mot repéré, mais on peut imaginer que les pauses seraient globalement bien moins longues s'il s'agissait juste de corriger une erreur qui a été déjà repérée en parallèle de l'écriture du mot, surtout lorsqu'il s'agit par exemple simplement d'un accent à ajouter. Cela conduit à penser que la modification

orthographique s'inscrit dans un temps de lecture et de réflexion plus large que la simple détection et correction d'erreur.

Si la clause est une « unité de production », il peut être attendu qu'à la fin de celle-ci il y ait de manière quasi-systématique un retour sur le texte, pour vérifier l'adéquation entre ce qui était conceptualisé et ce qui a été écrit. Or, la persistance de nombreuses erreurs d'orthographe met en doute cette relecture de surface systématique. Même dans le cas où une erreur est corrigée en fin d'unité, il est difficile de savoir si c'est suite à une relecture complète de la clause, ou parce que la correction d'une erreur détectée pendant l'écriture de la clause a été différée jusqu'à la complétion de celle-ci. Cependant, un des cas de « révision » traité dans cette partie peut appuyer l'hypothèse de relectures en fin de clause : les occurrences de « reprise esthétique ». Ces révisions ont la particularité de n'apporter aucune modification sur le mot : il peut s'agir de repasser une lettre, ou un diacritique, mais sans que la version originale nécessite une telle reprise. Logiquement, ces révisions ne devraient pas être couteuses à effectuer : or, près de 40% de ces reprises esthétiques sont précédées d'une pause « Q4 », alors qu'elles sont effectuées dans près de 80% des cas sur le mot précédent. Une explication peut être avancée si l'on regarde l'origine de ces révisions : de manière majoritaire (65%), ces reprises sont initiées en fin de clause, généralement après le signe de ponctuation. On peut alors supposer qu'en fin d'unité le scripteur entreprenne une relecture de ce qui a été produit ; pendant ce laps de temps, le scripteur peut, peut-être même sans s'en rendre compte, repasser sur le mot le plus proche, ou « ponctuer » la fin de sa relecture par une telle reprise esthétique. Outre la vérification de l'unité qui vient d'être écrite, cette relecture peut avoir pour but d'aider à planifier ce qui suit.

1.6.2. La formulation, une planification incrémentale ?

Dans le chapitre précédent, différents aspects de la formulation ont été analysés en lien avec les données chronométriques : ainsi, dans des séquences de texte où il n'y avait pas de correction, nous avons pu voir quels éléments étaient plus ou moins difficiles à planifier. L'analyse des modifications apportées au texte peut donner des indices sur ce qui peut être modifié, et ce qui est rarement remis en cause en cours d'écriture de la clause. Pour cela, le plus judicieux est d'observer les corrections apportées sur les mots. Celles-ci sont de plusieurs types, et sont plus ou moins « de surface » ou « de contenu », même si cette distinction n'est pas toujours claire (Lindgren & Sullivan, 2006a). Tout d'abord, on peut identifier la suppression de quelques morceaux de mots à cause d'inversions de lettres, ce qui s'apparenterait à ce qu'on appelle « fautes de frappe » pour les textes dactylographiés : par exemple « eper » supprimé et remplacé par « pressant » (S6), ou « cel » interrompu et remplacé par « clefs » (S26). Il arrive également que des mots aient été oubliés : « jusqu'à _ la^ (isser)[74]» interrompu

[74] Le « ^ » indique que le mot a été interrompu pour effectuer la révision à l'emplacement indiqué par « _ » ; ici, le scripteur finit ensuite d'écrire le mot pour obtenir le verbe « laisser ».

pour ajouter un « ne » (S9). Dans certains cas, on ne sait pas si le mot a été oublié, ou si la structure a été changée par rapport à ce qui était prévu : dans le cas de « regard si noir et intense que j » (S26) suivi de la suppression de « j » puis de l'écriture de « lui jetait le détenu », il est délicat de savoir s'il s'agit d'un oubli du pronom « lui », ou si le scripteur avait par exemple planifié de dire « jetait sur lui », puis a changé d'avis et opté pour la formulation « lui jetait ». Dans ces cas, les constituants de la clause changent peu, mais ils peuvent être réorganisés pour rendre la lecture plus fluide. Dans quelques cas, les suppressions de mots laissent apparaitre des modifications lexicales, sans que cela n'ait un impact sur le reste de la structure : ainsi, dans « Dimitri se sai », « se sai » est supprimé et remplacé par « s'empara » (S11). On peut fortement supposer que le scripteur avait commencé à écrire « se saisit », avant de remplacer ce verbe par un synonyme. Dans tous les cas mentionnés jusqu'ici, il apparait que la clause est globalement bien planifiée, mais que des petites modifications (lexicales, voire syntaxiques) peuvent être entreprises au cours de la rédaction. Il est très rare que le scripteur s'engage dans l'écriture d'une clause, puis en supprime une partie avant de la modifier profondément. Un seul cas de ce type émerge de nos données : en début de phrase, le scripteur écrit « Le », puis l'efface et écrit « Puis, il sortit » (S12). Dans ce cas, on peut supposer que la phrase a été entièrement reconceptualisée et reformulée après l'écriture du « Le » ; cela est appuyé par le fait que la pause avant l'écriture de « Puis » dure 32 secondes, ce qui est très long ; la pause avant « Le » durait 4,7 secondes, ce qui n'est pas particulièrement long pour une pause avant une nouvelle phrase. Dans l'ensemble de notre corpus, les scripteurs ne reviennent jamais a posteriori sur un mot pour le remplacer par un synonyme, ou changer sa nature (par exemple, remplacer un groupe nominal par un pronom), ce qui en théorie pourrait être possible pour améliorer la cohésion du texte. Cela est peut-être dû au médium d'écriture : bien qu'autorisés à faire des ratures, les scripteurs n'ont sans doute pas voulu en abuser, et ont donc évité les grandes reformulations. Cependant, on note trois cas de modifications portant sur le temps du verbe : « Dimitri est > était » (S7), « il se cacha > cachait » (S23), « il arrivait > arriva » (S23). Dans ces cas de changement de temps, l'enjeu dépasse le cadre syntaxique pour affecter le niveau discursif. En effet, le choix de mettre un présent « de narration » à la place de l'imparfait n'aurait pas été neutre, puisque cela aurait été un moyen pour le scripteur de rendre le récit plus « vivant » ; il a finalement opté pour la formule la moins marquée. Dans les cas d'hésitation entre l'imparfait et le passé simple (bien que dans le même texte, ces deux occurrences étaient éloignées l'une de l'autre), il s'agit ici pour le scripteur de distinguer ce qui relèvera du premier plan et du second plan. Même si ce n'est « que » une modification grammaticale, il s'agit donc plutôt dans ces cas d'une révision « de contenu » et non « de surface », puisque liée à un choix discursif.

1.6.3. La frontière entre unités A, difficile à matérialiser

Sur l'ensemble des révisions liées aux mots ou à la ponctuation, de nombreuses modifications concernent la question du lien entre unités Autonomes. Tout d'abord, il y a quelques cas où la frontière

des phrases (graphiques) a été redéfinie : dans deux cas la conjonction de coordination « ou » était remplacée par un point ; dans deux cas une virgule était remplacée par un point ; dans deux cas un point était supprimé, dont une fois où il était remplacé par la conjonction de coordination « et ». Chacune de ces révisions a été effectuée juste après l'écriture du premier mot de la nouvelle phrase ou clause. Sur l'ensemble du corpus, ces quelques cas sont donc plutôt rares. Il apparait que les adultes savent quelle sera l'étendue de la phrase, et reviennent rarement sur leur décision. Cependant, la longueur de certaines pauses en frontière d'unité A, marquée ou non par un point, peut justement s'expliquer par cette hésitation à clôturer la phrase.

Il y a également plusieurs révisions témoignant d'une hésitation quant à la meilleure marque de liaison entre deux unités Autonomes au sein d'une même phrase. Tout d'abord, il est arrivé qu'une marque ait été choisie pour finalement être remplacée : deux cas où la conjonction de coordination « et » a été remplacée par une virgule, et un cas où la conjonction « mais » a été remplacée par « et ». Enfin, il y a deux cas où une virgule a été insérée entre deux unités A, alors qu'il n'y avait aucune marque (l'absence de marque n'étant pas conforme aux normes typographiques). Là encore, les corrections ont été faites immédiatement après l'écriture du premier mot de la clause. Ces insertions de signe de ponctuation sont globalement assez rapides, car sans doute identifiés pendant l'écriture du premier mot de la clause. Cependant, on voit que le choix de la marque pour lier deux unités A n'est pas toujours automatisé lors de la production. Les résultats présentés dans le chapitre sur les données développementales montraient que les adolescents faisaient beaucoup d'écarts à la norme concernant la ponctuation entre unités A, et que cela restait la plus grande difficulté pour les élèves les plus âgés. Ici, nous voyons que les adultes connaissent la norme et savent repérer rapidement s'il manque une marque ou si celle-ci n'est pas adéquate, mais ils se placent pour cela dans la perspective du récepteur.

2. Étude de cas

2.1. Présentation du texte étudié

Nous avons vu que certaines hypothèses, concernant par exemple la gestion de l'orthographe, de la formulation ou de la segmentation au cours de l'écriture, pouvaient être éclairées par l'étude des révisions. Cependant, les cas de révisions étaient globalement rares dans le corpus « Dimitri », c'est pourquoi il avait été possible d'étudier les données chronométriques en écartant les séquences de révision. Pour clore ce chapitre, nous allons présenter une étude de cas qui montre que, dans certaines situations, l'analyse des révisions est beaucoup plus informative que celle des données uniquement chronométriques. Cette étude sera l'occasion d'expliciter une réflexion d'ordre méthodologique sur l'étude des processus rédactionnels à travers les empreintes et indices laissés sur le texte.

Ce texte a été obtenu dans le cadre de la collecte du corpus « Dimitri », il a donc été recueilli dans les mêmes conditions : il s'agissait d'une étudiante, qui avait pour consigne d'écrire la suite du récit *L'évasion* de manière compréhensible pour un enfant de 10 ans et avec le minimum d'erreurs d'orthographe ; le texte devait faire entre une demi-page et une page ; le temps d'écriture était libre, et il était possible de faire des corrections sur le texte. Le propos de cette étude est de montrer la complémentarité des outils à notre disposition : produit final, manuscrit, ordre des révisions, durées des pauses. Pour cela, nous allons volontairement procéder par étapes : nous débuterons avec la perspective du lecteur qui n'a devant lui que le produit final, puis dans la perspective du chercheur qui n'a à sa disposition que le manuscrit, puis dans celle du chercheur qui peut visionner le déroulement de l'écriture, et enfin dans celle du chercheur qui peut en outre avoir à sa disposition les données chronométriques liées à ce texte. Cette démarche reflète, à petite échelle, la manière dont la production écrite a pu être étudiée depuis la deuxième moitié du XXe siècle : après un intérêt pour l'écriture en tant que processus grâce à l'étude de manuscrits, l'utilisation d'enregistrement vidéos puis l'enregistrement de la rédaction grâce à des tablettes graphiques ou des ordinateurs a permis d'affiner les connaissances dans ce domaine. Chaque étape sera constituée de deux parties : des points généraux de description seront donnés[75], puis des hypothèses seront émises concernant certains aspects de l'écriture, le but étant de répondre à ces hypothèses au fur et à mesure de l'étude. En fin d'étude, une synthèse sera proposée sur les indices à disposition.

[75] Il arrivera que certaines données soient comparées à la « norme » du corpus « Dimitri », pour souligner les similitudes ou au contraire les différences entre cette production et celles étudiées précédemment.

2.2. Le produit final

La Figure 66 présente le produit final, dactylographié à partir du manuscrit du scripteur. Il serait trop long et pas forcément pertinent de faire une analyse linguistique exhaustive de ce texte, mais il est possible de faire quelques remarques sur la forme (orthographe, syntaxe) et le fond (scénario, champs lexicaux) du texte.

> Leste comme un chat, il se faufila entre les jambes du gardien et disparut aussitôt, avant que ce dernier, stupéfait, n'ait pu faire le moindre geste pour le retenir.
>
> Dimitri, le cœur battant, se dissimula derrière la plante verte qu'il avait repérée le jour même de son arrestation. Retenant son souffle, il attendit quelques instants derrière les larges feuilles et vit le gardien furieux passer devant lui sans le remarquer. Il aurait préféré duper Norbert, l'affreux gardien du matin, plutôt que celui-ci, dont il ne connaissait pas le nom, mais qui avait toujours fait preuve d'amabilité envers Dimitri. Mais l'heure n'était pas aux états d'âme, il fallait fuir. Peu importaient les conséquences, tant qu'il parvenait à sortir d'ici. Il allait enfin retrouver ses parents, sa chambre et sa tranquillité. Lorsqu'il fut certain d'être seul dans le couloir complètement silencieux, il sortit de sa cachette et descendit par l'escalier de service à pas de loup. Il jeta un œil par la vitre qui donnait sur le hall d'entrée : le veilleur de nuit était là, assis dans son fauteuil à roulettes, presque assoupi. Dimitri avait bon espoir qu'il s'endorme, il attendrait aussi longtemps que nécessaire. Il voyait l'homme lutter contre le sommeil, papillotant des yeux. Mais au moment où il allait s'endormir, le téléphone sonna. « Comment ? Le petit Dimitri ? Il a disparu ? Mais ses parents viennent le chercher demain matin ! Oui, le docteur Bécot m'a dit que le traitement s'était bien passé et qu'il pouvait rentrer chez lui. Oui oui, toute la famille sera là, même son grand frère. Mais c'est une surprise ! » En entendant ces mots, Dimitri écarquilla les yeux et s'aperçut qu'il n'y avait ni mirador, ni barreaux, ni cellule. Et il regagna sa chambre en courant.

Figure 66. Texte du scripteur S31 - Produit final

2.2.1. Description générale

La version finale de ce texte compte 361 séquences, réparties en 21 phrases. En moyenne, une phrase comporte 1,33 unités A, 2,38 clauses et 17,19 séquences. Il y 22 clauses subordonnées sur un total de 50 clauses, soit 44% de clauses subordonnées. On note la présence de 4 subordonnées circonstancielles situées en début de phrase, ainsi que 7 circonstancielles avec un verbe non fléchi (participe présent ou passé). Dans les cas d'unités A coordonnées au sein d'une même phrase, il y a toujours une ellipse du sujet de la 2ème unité A lorsque celui-ci est coréférentiel à celui de l'unité A précédente. Dans l'ensemble, ce texte ne s'écarte pas fortement des moyennes trouvées sur le corpus « Dimitri » : un certain nombre de structures complexes sont utilisées, témoignant d'un usage diversifié des structures syntaxiques du français, sans que le texte ne devienne trop difficile pour la compréhension. Par ailleurs, aucune erreur d'orthographe n'a été relevée dans le produit final.

Si le texte ne se distingue pas par l'organisation syntaxique, il est en revanche très original d'un point de vue du scénario. En effet, le titre du récit à finir était « L'évasion ». Par conséquent, les scripteurs du corpus ont généralement activé un schéma narratif correspondant à un récit d'évasion : maitrise du gardien, fuite dans les couloirs, sortie de la prison ; ces parties étant plus ou moins développées suivant les récits. Certes, quelques textes étaient un peu plus originaux : ainsi, le gardien était parfois complice ; parfois, l'ensemble du texte résidait dans les pensées de Dimitri au moment où il est confronté au gardien dans la cellule ; enfin, dans deux textes on apprenait que toute l'histoire n'était qu'un rêve, puisqu'à la fin Dimitri se réveillait dans sa cellule. En revanche, le texte du scripteur S31 est beaucoup plus original : il apparait que Dimitri est en réalité un enfant, et que l'action se déroule dans un établissement médical. Ce dénouement apparait progressivement aux yeux du lecteur :

- il y a tout d'abord l'indication « retrouver ses parents », qui peut suggérer que Dimitri est jeune, sans que cela ne soit nécessaire ;
- cette hypothèse est confirmée par la désignation « petit Dimitri » ;
- il est ensuite question de « docteur » et « traitement » ;
- enfin, l'avant-dernière phrase confirme qu'il n'y pas de prison.

Ces changements progressifs sont illustrés par la succession des champs lexicaux : au début il y a un champ lexical lié à la prison et à la fuite (« gardien », « arrestation », « se faufiler », « disparut », « fuir », « sortir », « cachette », « veilleur de nuit ») ; puis on trouve les termes liés au bâtiment, qui sont plutôt neutres et pourraient s'appliquer à différents types d'établissement (« plante verte », « couloir », « escalier de service », « vitre », « hall d'entrée », « téléphone », « chambre ») ; ce n'est qu'à la fin qu'apparaissent les termes relevant du domaine médical (« docteur », « traitement », « rentrer chez lui »). En outre, plusieurs termes soulignent l'importance du fait qu'il s'agit d'un petit garçon : « parents », « petit », « grand frère », « famille ».

Un des points originaux de ce récit est que la découverte est progressive pour le lecteur, mais également pour le personnage de l'histoire : ce n'est qu'à la fin du récit que l'enfant « s'aperçut qu'il n'y avait ni mirador, ni barreaux, ni cellule ». Cette indentification entre le lecteur et le personnage est renforcée par le point de vue interne adopté par le narrateur au long de l'histoire. En effet, durant tout le récit il est décrit ce que le personnage principal voit (« qu'il avait repéré », « vit », « voyait »), entend (« en entendant ces mots »), connait (« ne connaissait pas le nom », « fut certain », « s'aperçut que »), souhaite (« aurait préféré », « il fallait », « peu importaient », « avait bon espoir ») et compte faire (« il attendrait »). Certains termes marquent également la subjectivité, comme l'adjectif placé avant le nom dans « affreux gardien » ou l'adverbe « allait enfin » soulignant l'attente du personnage. Ainsi, le « Dimitri enfant » se dévoile peu à peu aux yeux du personnage, ainsi qu'aux yeux du lecteur. Certes, la découverte pour le personnage est moins grande que pour le lecteur, puisque logiquement celui-ci sait dès le début qu'il est un enfant, et découvre simplement qu'il ne se trouve pas dans une prison. Le lecteur comprend à la fin, tout comme le personnage, que la prison, le mirador et la cellule ne sont que des

éléments imaginés par l'enfant, qui s'est sans doute construit un univers carcéral autour de lui. Il est alors possible de relire le texte à la lumière de cette information : si la cellule devient sa « chambre », le lecteur suppose alors que le « gardien » est sans doute un infirmier. Un terme peut sembler cependant étrange : l'« arrestation », lexème fortement lié au champ lexical de l'incarcération ; la seule hypothèse est d'imaginer que cette « arrestation » renvoie au jour où Dimitri a été admis dans cet établissement médical, ce qui a peut-être été vécu comme un emprisonnement.

2.2.2. Questions soulevées et premières hypothèses

Question 1 : L'analyse de la « forme » du texte fait ressortir que celle-ci est conforme aux attentes de la consigne : il n'y a pas d'erreurs d'orthographe, et le texte est relativement complexe d'un point de vue syntaxique tout en restant compréhensible pour un enfant. On peut alors se poser la question : dans quelle mesure la correction syntaxique et orthographique de ce texte a-t-elle été retravaillée par le scripteur ?

Hypothèse liée à la question 1 : L'absence d'erreurs d'orthographe et la cohérence et cohésion observée entre les phrases suggère que des relectures ont été effectuées afin d'éviter tout écart à la norme.

Question 2 : Concernant le scénario, il apparait que la « vraie » histoire se dévoile progressivement au personnage et au lecteur. La question que l'on peut se poser est : se dévoile-t-elle également au scripteur ? Autrement dit, le scripteur avait-il dès le début l'idée de ce scénario, du moins dans ses composantes principales, ou celui-ci a-t-il évolué en cours d'écriture ?

Hypothèse liée à la question 2 : À ce stade, il n'est pas possible d'avoir des certitudes sur ce point : le fait que le début de l'histoire ne soit pas très original (champs lexicaux de la prison, de la fuite et des bâtiments) peut-être dû à une première idée de scénario « classique », tout comme à une volonté délibérée de la part d'induire en erreur le lecteur. En outre, l'ensemble est cohérent, mais il est possible que le scripteur ait apporté des modifications sur son texte pour assurer cette cohérence.

Ces questions portent donc d'une part sur la formulation (Q1) et d'autre part sur la macro-planification (Q2).

2.3. Le manuscrit

La Figure 67 est une copie de la version manuscrite du texte étudié. Contrairement aux autres textes du corpus, beaucoup de ratures sont visibles, l'étude de cette version manuscrite donnera donc certainement des informations supplémentaires par rapport au produit final.

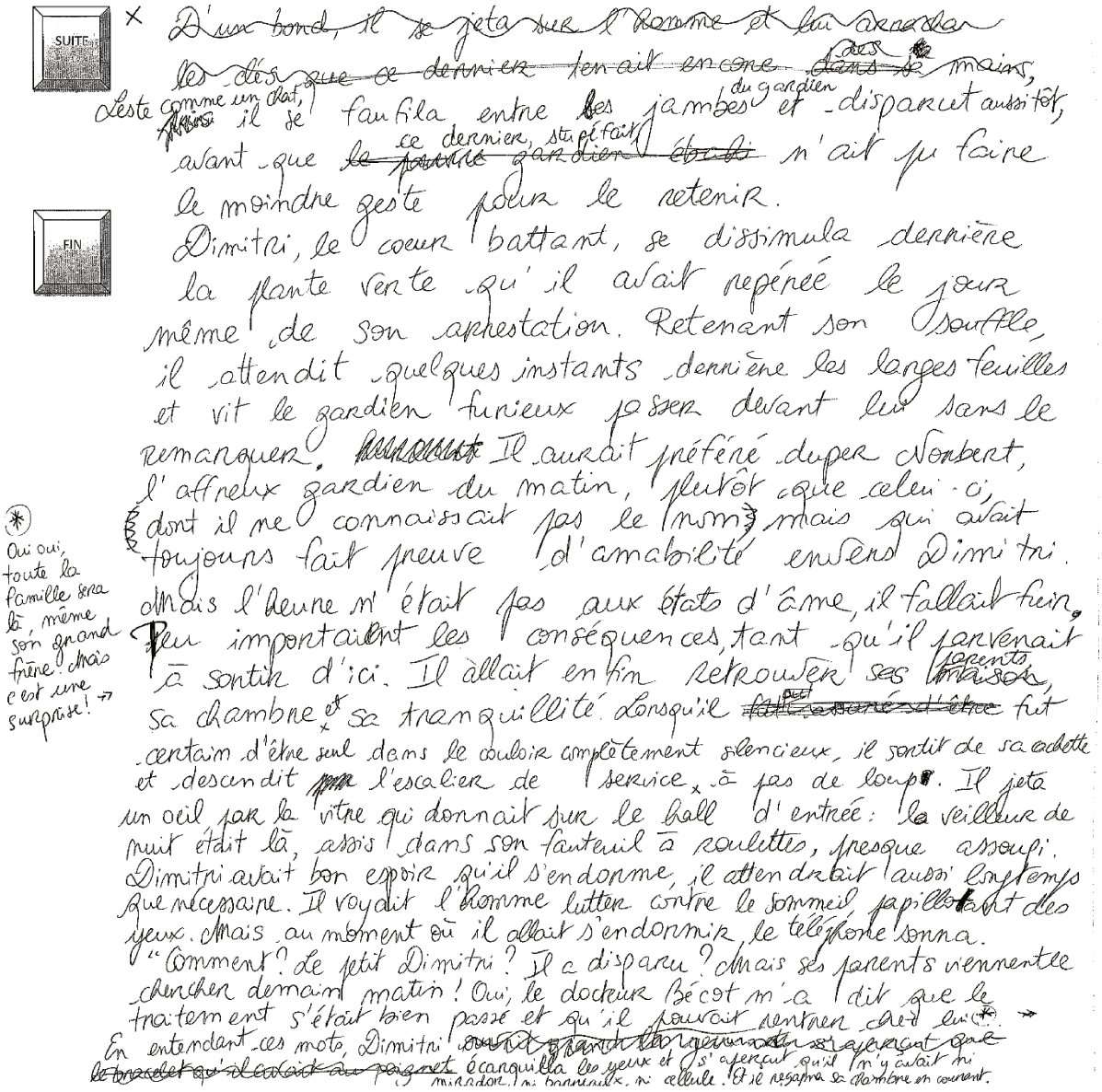


Figure 67. Texte du scripteur S31 - Manuscrit

2.3.1. Description générale

Dans ce manuscrit, il est possible d'identifier différentes opérations de révision (ajouts, suppressions, remplacements, modification orthographique) portant sur différentes unités (groupes de mots, mots, ponctuation). Ces révisions sont relevées dans le Tableau 66.

Opération	Taille du segment affecté	Séquence
Ajout	groupe de mots	Leste comme un chat
		du gardien
		Oui oui, toute la famille sera là, même son grand frère. Mais c'est une surprise ! »
Suppression	groupe de mots	D'un bond, il se jeta sur l'homme et lui arracha les clés que ce dernier tenait encore dans sa main
		fut assuré d'être
		ouvrit grands les yeux et s'aperçut que le bracelet qu'il avait au poignet
	mot	des
		la
		puis
		hurlant
		put
		par
	ponctuation	(
		)
		,
Remplacement	groupe de mots	le pauvre gardien ébahi > ce dernier stupéfait
	mot	ses > les
		maison > parents
	pct > mot	, > et
	ponctuation	. > ,
	majuscule	peu > Peu
Modification orthographique		loups > loup
		papillottant > papillotant

Tableau 66. Révisions identifiées dans le manuscrit

2.3.2. Exploration des questions soulevées

2.3.2.1. La présence de reformulations (Q1)

Si l'on va au-delà de la simple énumération d'opérations isolées, on peut identifier les changements effectués par le scripteur. Nous nous étions interrogés sur d'éventuelles modifications liées aux structures syntaxiques et à l'orthographe.

Tout d'abord, deux corrections orthographiques sont identifiables : « à pas de loups > à pas de loup » et « papillottant > papillotant » (les deux « t » ont été transformés en une seule lettre, qui apparait plus épaisse). Cependant, nous ne savons pas quand elles ont été effectuées.

Nous observons ensuite des changements concernant la ponctuation. Une seule modification concerne la segmentation en phrases : « il fallait fuir, peu importaient > il fallait fuir. Peu importaient ». Les autres hésitations concernent la ponctuation intraphrastique, que ce soit pour délimiter une relative non restrictive (suppression des parenthèses), finir une énumération (« sa chambre, sa tranquillité > sa chambre et sa tranquillité ») ou lier un complément de manière (« descendit l'escalier de service à pas de loup »). Concernant la délimitation de la clause « dont il ne connaissait pas le nom », on peut remarquer que le fait de garder les parenthèses telles qu'elles étaient inscrites aurait été erroné : les parenthèses indiquent que le segment pourrait ne pas être lu, or il ne serait pas possible de lire uniquement « plutôt que celui-ci, mais qui avait toujours fait preuve d'amabilité envers Dimitri ». Une hypothèse est que le scripteur a mis des parenthèses, mais en continuant sa phrase a vu qu'il aurait dû ne pas refermer la parenthèse, et les a donc supprimées.

D'autres changements visibles concernent des choix lexicaux. Dans certains cas, il s'agit de remplacements d'un mot par un synonyme : « ébahi » > « stupéfait », « fut assuré » > « fut certain », ou d'une hésitation entre deux formes possibles et quasiment synonymes (« descendit par l'escalier » > « descendit l'escalier »). D'autres changements concernent la forme lexicale, mais ils ont surtout un impact discursif, puisqu'ils concernent la manière dont un référent est mentionné au fur et à mesure du texte : « ses jambes > « les jambes du gardien » et « le pauvre gardien > « ce dernier ». Cela témoigne d'une attention portée à la cohésion du texte et à la meilleure formulation des référents. Ces deux derniers changements se situent dans les premières lignes du texte, qui ont fait l'objet de nombreuses révisions. L'attention portée à ce passage suggère que ces révisions se sont effectuées en plusieurs étapes. En effet, les reformulations lexicales mentionnées, auxquelles il faut ajouter des hésitations raturées avec les mots « dans », « des », « sa », « la », témoignent d'un tâtonnement lexical qui cherche sans doute à s'accorder au mieux avec ce qui précède : une rature d'un passage peut entrainer une reformulation d'un autre groupe de mots. Le manuscrit fait ressortir des types de rature différents, suggérant plusieurs étapes de suppressions. Par exemple, même si les deux premières lignes sont entièrement raturées, il semble que quatre segments raturés ressortent : « D'un bond, il se jeta sur l'homme et lui arracha les clés des mains », « que ce dernier tenait encore », « dans sa », « la ». Ainsi, il semble que le segment « que ce dernier tenait encore » ait été supprimé antérieurement aux autres révisions. De même, l'adjectif « pauvre » n'est pas raturé de la même manière que « le gardien ébahi » : il est possible qu'il ait été supprimé avant les autres mots, ce qui serait davantage cohérent avec le remplacement de « le gardien ébahi » par « ce dernier stupéfait », davantage équivalent d'un point de vue sémantique.

Par rapport aux points soulevés à partir du produit final, l'analyse du manuscrit suggère que le texte a été à plusieurs reprises retravaillé pour des éléments tant de « surface » (orthographe, synonymes) que de contenu (suppression de phrases entières, continuité référentielle). La question qui se pose concernant ces reformulations est maintenant : en quoi les modifications apportées amélioraient-elles le texte produit ?

2.3.2.2. Évolution du scénario (Q2)

Notre deuxième interrogation portait sur un changement potentiel de scénario en cours d'écriture, qui n'était pas certain si l'on ne s'en tenait qu'au produit final. L'étude du manuscrit semble appuyer l'hypothèse du changement de scénario. En effet, le premier indice pour le lecteur que Dimitri était jeune était le mot « parents », présent dès la partie sur la fuite dans le couloir. Or, il apparait que ce mot est présent suite à une reformulation, remplaçant le nom « maison », qui aurait pu convenir à un individu de tout âge. On ne sait cependant pas à quel moment ce mot a été modifié : est-ce plus tard, lorsque le scripteur a eu l'idée de présenter un Dimitri enfant, ou est-ce un choix délibéré à la fin de la rédaction pour donner un indice au lecteur, étant donné que le mot « maison » aurait pu être conservé sans nuire à la cohérence de l'ensemble ? Cette deuxième possibilité pourrait alors être cohérente avec l'hypothèse d'une planification du scénario dès le début. Il est en tout cas certain que, « parents » n'étant pas présent dès le premier jet, le premier indice concernant l'identité de Dimitri est la mention « petit Dimitri » : à partir de là, on est certain que le scripteur a en tête ce scénario. Le fait que le scripteur ait manqué de place en fin de rédaction suggère que le scénario n'était pas entièrement établi dès le début de la rédaction : en effet, la consigne étant de rédiger entre une page et une demi-page, le scripteur aurait sans doute mieux gérer l'espace s'il avait prévu ce changement de situation et les longs dialogues. D'ailleurs, l'ajout d'une phrase entière du dialogue dans la marge témoigne que celui-ci n'avait pas été totalement planifié. Enfin, ce « changement de scénario » probable nous amène à reconsidérer les révisions du début de texte. La suppression de la première phrase évite de présenter un Dimitri agressif (« arracha ») et de taille à affronter le gardien ; à l'inverse, l'apposition « leste comme un chat » présente un personnage agile et en mesure de se faufiler, ce qui est cohérent avec un personnage enfant. On peut donc faire l'hypothèse que ces changements ont été effectués tardivement, en vue de rendre l'ensemble du texte plus cohérent.

Ainsi, il semble que le scénario ait été modifié en cours d'écriture. La question qui se pose alors est : quand ce changement a-t-il eu lieu, et quels éléments ont été modifiés pour rendre l'ensemble du texte cohérent avec le nouveau scénario ?

2.4. Étude du « film de l'écriture »

La rédaction du texte a été enregistrée grâce au logiciel à une tablette graphique couplée au logiciel Eye & Pen : il ne s'agit donc pas d'un enregistrement vidéo, mais d'un enregistrement des données qui

permet de reconstruire la production dans ses aspects temporel et spatial. Dans cette partie, de nouveaux indices seront donnés grâce à l'étude des séquences de révision. Nous verrons tout d'abord quelles informations complémentaires sont apportées grâce au film de l'écriture, puis nous tenterons de répondre aux questions issues de la partie précédente.

2.4.1. Description générale

2.4.1.1. Affinement du relevé des révisions

L'enregistrement permet d'identifier des révisions qui n'étaient pas visibles sur le manuscrit. Ainsi, l'adverbe « aussitôt » a été ajouté a posteriori ; un retour sur le manuscrit montre en effet qu'il est écrit plus petit, mais cela aurait pu être dû à des contraintes d'espace. Les autres modifications qui étaient non visibles concernent la ponctuation : un ajout d'une virgule après « celui-ci », un point remplacé par un point d'exclamation après « surprise ». Il apparait également deux hésitations concernant la segmentation en phrase : la virgule après « remarquer » a été modifiée en un point ; à l'inverse, le point après « conséquences » a été remplacé par une virgule. Enfin, l'enregistrement montre que les parenthèses, sur lesquelles nous nous étions interrogés précédemment, ont été ajoutées a posteriori, puis supprimées.

L'enregistrement est également le moyen de relever les cas de « reprise esthétique » : celles-ci n'apportant par définition aucune modification au mot, il était normal de ne pas pouvoir les identifier sur le manuscrit. Or, cette production compte 43 reprises esthétiques, soit 48% des révisions de ce texte. À partir des informations fournies par l'enregistrement, il est possible de décrire avec précision le nombre de révisions du corpus, comme illustré par la Figure 68. Ce texte compte au total 90 séquences de révision, soit 63% du corpus « Dimitri », qui était constitué de 30 textes. Des différences sont notables entre les révisions relevées dans ce texte et dans le corpus, ce qui souligne la difficulté de généraliser certains résultats. Comparant le texte du scripteur S31 au corpus « Dimitri », la Figure 69 illustre la répartition des révisions en fonction du type de révision, et la Figure 70 la distribution des révisions en fonction du type de segment révisé. Concernant le type d'opérations, le texte du scripteur S31 comporte en proportion beaucoup moins de modifications orthographiques, mais davantage de reprises esthétiques, d'ajouts et de remplacements. En outre, dans une proportion non négligeable (16%) ces révisions concernent des groupes de mots, alors qu'il n'y avait pas de telles occurrences dans le corpus. Il arrive par ailleurs que des reprises esthétiques portent sur des ratures.

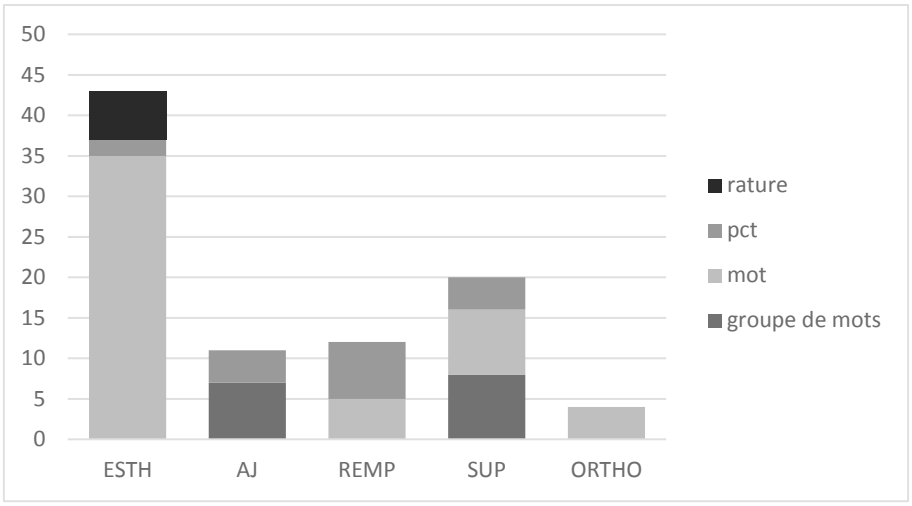


Figure 68. Nombre de révisions du texte du scripteur S31 en fonction du type d'opération et de segment révisé

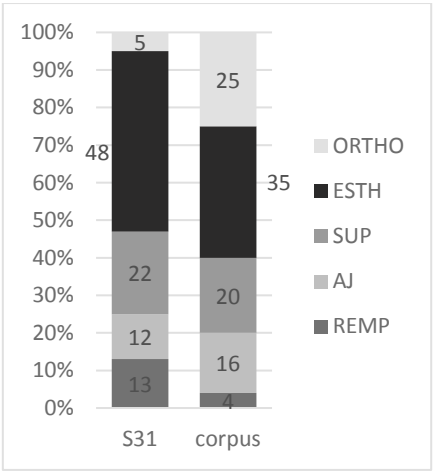


Figure 69. Distribution des révisions du texte du scripteur S31 et du corpus en fonction du type d'opération (en %)

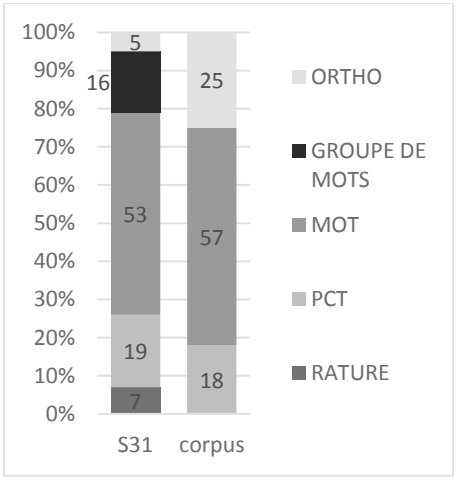


Figure 70. Distribution des révisions du texte du scripteur S31 et du corpus en fonction du type de segment révisé (en %)

Par ailleurs, l'enregistrement a bien confirmé la complexité de certaines séquences de révision. Ainsi, les hypothèses formulées à partir des différents types de rature du premier paragraphe sont confirmées. En outre, le film de l'écriture montre également la complexité de l'écriture des dernières lignes, où il y a une succession de ratures et d'ajouts.

2.4.1.2. Origine des révisions

L'enregistrement du texte permet également de savoir quand les révisions ont été initiées. Tout d'abord, il peut être remarqué que, sur les 90 séquences de révisions, seules 35 interrompent l'écriture du texte. Pour les autres cas, les révisions s'enchainent les unes aux autres au cours de « sessions » de révision, comme ce qui avait été relevé par Severinson Eklundh et Kollberg (2003).

Les origines des 35 révisions interrompant l'écriture linéaire du texte ont été à nouveau réparties en « longues », « moyennes », « courtes », « mot précédent » et « mot en cours ». De nouveau, des grandes différences illustrées par la Figure 71 peuvent être soulevée : par rapport au corpus, les révisions

immédiates (n et n-1) sont beaucoup moins fréquentes dans le texte du scripteur S31, qui privilégie les révisions courtes et moyennes.

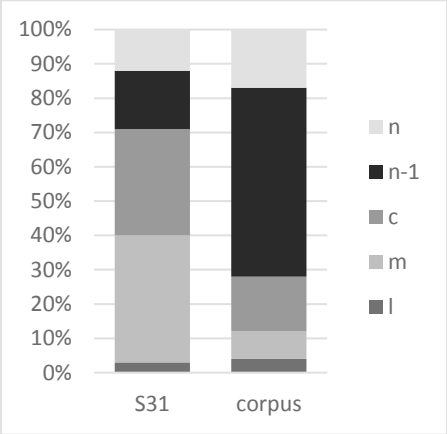


Figure 71. Distance entre l'origine et le site de la révision - texte du scripteur S31 et corpus (répartition en %)

Il est possible de s'interroger sur le lieu où sont initiées ces révisions. Dans ce texte, il est frappant de constater que 13 de ces révisions (37%) sont initiées en fin de ligne, ce qui suggère un comportement particulier lié à l'aspect spatial de l'écriture. Ces révisions initiées en fin de ligne correspondent parfois à des fins d'unités linguistiques : 5 cas de fin de phrases, 2 cas de fin d'unité A, et 3 cas où l'unité A aurait pu être terminée tout en étant syntaxiquement correcte. Dans 12 autres cas, la révision est également initiée en fin d'unité linguistique, que ce soit en fin de phrase (7), fin d'unité A (1), fin de clause (1) ou fin potentielle d'unité A. Le fait de relever la segmentation en unités syntaxiques non pas seulement à partir du produit fini, mais plutôt en fonction de ce qui a été écrit jusque-là, s'inspire de la segmentation proposée par Spelman-Miller (2002a, 2002b), qui indique des frontière par rapport au « point de complétion potentiel ». En effet, il n'est pas exclu qu'un scripteur ayant tendance à s'arrêter en fin d'unité linguistique fasse un arrêt lorsqu'il estime avoir fini son unité, effectue des révisions, et finalement ajoute une suite à son unité. Cela soulève la question de l'unité de production, donc telle que prévue par le scripteur, qui peut différer de l'unité perçue par le lecteur lorsqu'il a accès au produit final. Sur l'ensemble de ce texte, moins de 30% des révisions ne sont donc pas initiées à une frontière linguistique ou en fin de ligne. Ce nombre n'est cependant pas négligeable, et les analyses des révisions elles-mêmes nous renseigneront peut-être sur ces cas.

Dans le texte, 15 sessions de révisions sont identifiables, c'est-à-dire des moments pendant lesquels se succèdent plusieurs révisions. Dans 80% des cas, les révisions sont spatialement proches les unes des autres au sein d'une même session de révision, ce qui signifie qu'un passage en particulier est retravaillé. Dans 13% des cas, la session de révision se termine par une révision proche du point d'origine de la révision, c'est-à-dire que la révision effectuée amène le scripteur à modifier ce qu'il était en train d'écrire, même si les séquences sont éloignées dans le texte. Enfin, il n'y a que 7% de cas où une révision est effectuée loin de la séquence précédente et loin de la séquence d'origine ; cela concerne notamment les corrections effectuées en fin de texte, qui se situent à différents endroits.

2.4.2. Exploration des questions soulevées

La première question portait sur le déroulement et l'intérêt des reformulations, avec notamment des interrogations sur les premières lignes du texte, qui avaient fait l'objet de nombreuses modifications et semblaient particulièrement intéressantes à étudier. En effet, elles révélaient une attention particulière au maintien d'une cohésion lexicale appropriée. Nous avions également émis l'hypothèse que des éléments avaient été supprimés ou ajoutés pour rendre le texte plus cohérent avec le scénario d'un Dimitri enfant. C'est donc ce passage qui sera particulièrement analysé à partir des données de l'enregistrement pour la question des reformulations. Nous traiterons la seconde question, portant sur l'évolution du scénario, à partir d'indices issus de différentes parties du texte.

2.4.2.1. Étapes de reformulation du début du texte (Q1)

L'enregistrement révèle que la réécriture de ce passage a été faite au cours de trois sessions de révisions. Nous allons donc expliciter ces trois étapes.

- Étape 1 (début 2'27, durée 54 secondes)

Cette session de révision débute au bout de 2'27 d'écriture, et dure 57 secondes. Elle est initiée à la fin de la 2ème ligne ; à ce moment de l'écriture, un point clôture la phrase se terminant par le mot « main ». Les opérations de révision sont précisément décrites dans le Tableau 67.

AJ – groupe de mots	de la
SUP – groupe de mots	dans sa
SUP – groupe de mots	que ce dernier tenait encore
REMP – mot	de > des
REMP – mot	main > mains
SUP – mot	la
ESTH	.
ESTH	rature "la"
ESTH	rature "que ce dernier tenait encore"
REMP – pct	. > ,

Tableau 67. Révisions en début du texte - Détails de l'étape 1

Au-delà de ces opérations isolées, les reformulations successives peuvent se résumer ainsi[76] :

- D'un bond, il se jeta sur l'homme et lui arracha les clés que ce dernier tenait encore dans sa main.
- D'un bond, il se jeta sur l'homme et lui arracha les clés que ce dernier tenait encore **de la** main.

[76] La nouvelle formulation est indiquée en gras.

- D'un bond, il se jeta sur l'homme et lui arracha les clés _ **des** main**s**.
- D'un bond, il se jeta sur l'homme et lui arracha les clés des main**s,**

À première vue, on pourrait penser que l'interrogation portait sur la préposition et le déterminant précédant « main » : « dans sa main » > « de la main » > « des mains ». En réalité, c'est la forme verbale elle-même qui est en jeu dans cette recherche : « tenir dans » > « tenir de » > « arracher de ». C'est donc le prédicat complet de cette phrase qui a été modifié, même si la trace est laissée simplement sur le groupe prépositionnel final. En outre, ce changement assez important, avec notamment la suppression d'une clause, conduit le scripteur à redéfinir sa segmentation, et à poursuivre sa phrase plutôt qu'un débuter une nouvelle. Ce remplacement est précédé de plusieurs reprises esthétiques sur le point final et les ratures précédentes.

- Étape 2 (début 10'58, durée 34 secondes)

Cette étape est initiée après 10'58 d'écriture, et dure 34 secondes. Elle intervient après l'écriture du participe présent « hurlant ». Elle débute sur des reprises esthétiques sur le mot « disparut » et la conjonction « et » qui le précède. Ensuite, le scripteur ajoute « aussitôt, » et supprime l'adjectif « pauvre » dans le groupe nominal « le pauvre gardien ». On peut être surpris du caractère hétéroclite de cette session de révision, puisqu'il y a des reprises esthétiques, ajouts, suppressions ; les trois premières opérations se situent sur le même syntagme, mais l'adjectif supprimé « pauvre » n'est pas directement lié à ce syntagme. Suite à ces révisions, qui portaient sur le début du texte, le scripteur modifie également ce qu'il était en train d'écrire, puisqu'il efface le participe présent et décide de débuter une nouvelle phrase.

- Étape 3 (début 20'23, durée 2 minutes)

Cette dernière session de révision sur le début du texte intervient au bout de 20'23 et dure deux minutes. Cette longue session interrompt un groupe nominal : elle est initiée entre « sa » et « tranquillité ». Les reformulations successives de ce passage sont les suivantes :

- D'un bond, il se jeta sur l'homme et lui arracha les clés des mains, puis il se faufila entre ses jambes et disparut aussitôt, avant que le gardien ébahi n'ait pu faire le moindre geste pour le retenir.
- ~~**D'un bond, il se jeta sur l'homme et lui arracha les clés des mains, puis**~~ il se faufila entre ses jambes et disparut aussitôt, avant que le gardien ébahi n'ait pu faire le moindre geste pour le retenir.
- il se faufila entre **les** jambes **du gardien** et disparut aussitôt, avant que le gardien ébahi n'ait pu faire le moindre geste pour le retenir.
- il se faufila entre les jambes du gardien et disparut aussitôt, avant que **ce dernier, stupéfait,** n'ait pu faire le moindre geste pour le retenir.
- **Leste comme un chat,** il se faufila entre les jambes du gardien et disparut aussitôt, avant que ce dernier, stupéfait, n'ait pu faire le moindre geste pour le retenir.

Conformément à une de nos hypothèses, il apparait bien que les reformulations « ses jambes > les jambes du gardien » et « le gardien > ce dernier » répondent à un besoin de cohésion lexicale concernant le référent « gardien ». En effet, dans la première version la chaine référentielle est : « l'homme / lui / ses / le gardien ». La suppression de la première phrase conduisait à la chaine référentielle « ses / le gardien ». Il était nécessaire de réintroduire plus explicitement le référent, d'où la mention « du gardien » dans la nouvelle phrase initiale du texte. La chaine « du gardien / le gardien » étant trop redondante, le scripteur a remplacé la dernière mention par « ce dernier ». On peut d'ailleurs remarquer que cette expression avait été utilisée dans la clause de la première phrase qui avait était raturée dès le début ; elle était donc relativement accessible, que ce soit en mémoire ou grâce au texte encore visible malgré la rature. Avant l'insertion du groupe adjectival « Leste comme un chat » en début de texte, le scripteur produit trois révisions esthétiques (sur les mots « ce », « stupéfait » et la rature de « puis »).

Contrairement à une de nos hypothèses, la reformulation de ce début de texte n'a pas été faite à la fin du texte, dans un souci de cohérence globale. Il est incertain, à ce stade, que le scripteur ait déjà l'idée de Dimitri enfant. Ces modifications donnent tout de même l'impression d'un personnage moins agressif, ce qui trouve d'ailleurs écho avec le terme « tranquillité », qui vient compléter le groupe nominal interrompu. On peut supposer que le scripteur souhaitait avoir une énumération à trois éléments, pour des considérations stylistiques (« sa maison, sa chambre, sa … »), mais sans avoir trouvé le troisième terme. Il s'est donc interrompu pour s'inspirer du texte déjà écrit, et a à cette occasion apporté des modifications au début du texte, nuançant le caractère du personnage ; le scripteur a ensuite pu terminer son énumération.

2.4.2.2. Évolution du scénario (Q2)

L'étude du manuscrit suggérait une modification du scénario en cours d'écriture grâce à deux indices : le remplacement de « sa maison » par « ses parents », et les modifications du premier passage pour rendre l'ensemble plus cohérent. Nous venons de voir que les modifications du premier passage avaient été faites assez tôt dans le texte, du moins bien avant le remplacement de « sa maison » par « ses parents ». Ce remplacement a été effectué après la clause « le téléphone sonna », soit juste avant la mention « petit Dimitri ». On peut alors faire l'hypothèse que l'idée que Dimitri soit un enfant est apparue tardivement au scripteur, peut-être seulement avant le discours direct rapportant l'appel téléphonique. Cependant, on ne peut pas exclure que l'idée soit arrivée progressivement, et les changements du début du texte parlant d'un personnage leste, peu agressif, ont pu aider à construire la personnalité du personnage jusqu'à en faire un enfant. Aucune modification majeure n'a été apportée à la fin de la production pour rendre l'ensemble davantage cohérent.

2.5. Étude des pauses

2.5.1. Description générale

Dans ce texte, les pauses en début de mot durent entre 52 ms et 87 secondes. Le Tableau 68 indique les seuils des quartiles et de déciles du texte du scripteur S31, en les mettant en regard avec les valeurs du corpus. Les pauses initiales de ce texte sont particulièrement longues : pour chaque seuil, la valeur de celui du texte étudié est plus élevée que la valeur maximum relevée dans le corpus, à l'exception du seuil « D10 », délimitant les 10% de pauses les plus longues.

	Q2	Q3	Q4	D9	D10
Texte – S31	492 ms	1236 ms	4568 ms	7103 ms	13722 ms
Corpus : moyenne	231,65 ms	357,03 ms	821,31 ms	1266 ms	4452 ms
Corpus : maximum	368 ms	608 ms	2714 ms	3804 ms	15154 ms
Corpus : minimum	116 ms	192 ms	356 ms	424 ms	840 ms

Tableau 68. Seuil des quartiles et déciles - Texte du scripteur S31 et corpus (en ms)

Si l'on adopte une approche linguistique en décrivant la durée des pauses en fonction de l'emplacement syntaxique, comme illustré par la Figure 72, on remarque que l'« effet de l'emplacement » apparait moins nettement que dans le corpus. La distinction entre phrases, unités A et clauses apparait essentiellement dans le quartile Q4 hormis les pauses D10 (c'est-à-dire la catégorie 75-90%) : en termes de pourcentage, il y a davantage de pauses « P » que de pauses « A », et davantage de pauses « A » que de pauses « C » ; on voit cependant peu de différence entre les pauses « C » et « I ». Le pattern inverse se retrouve pour les pauses plus courtes (Q1 à Q3). En revanche, les 10% de pauses les plus longues, supérieures à 13,7 secondes, sont moins sensibles à l'effet de l'emplacement syntaxique.

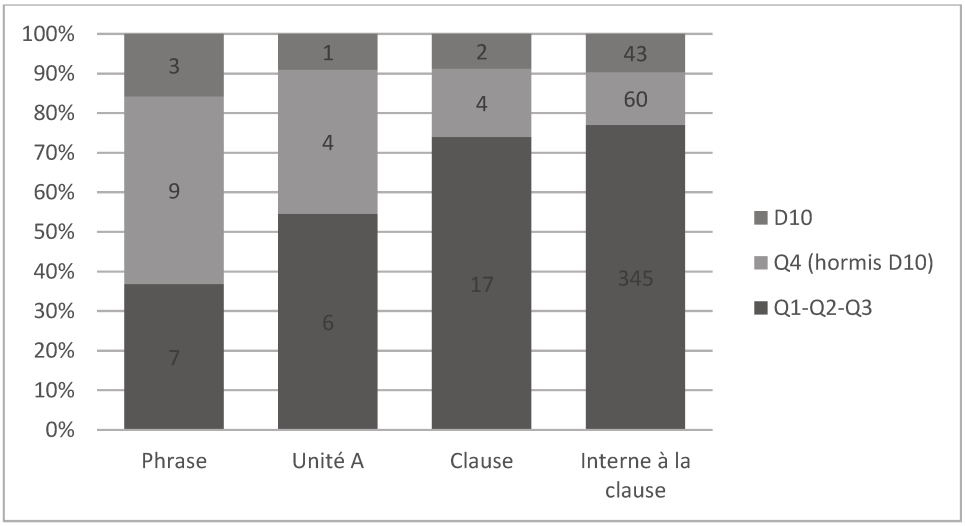


Figure 72. Répartition des pauses initiales en fonction de l'emplacement syntaxique (durées en quartiles et déciles)

On peut alors adopter une approche chronométrique afin de voir où se situent ces 10% de pauses les plus longues. La Figure 73 montre que presque la moitié des pauses D10 du texte correspondent à des révisions ; un quart sont situées à une frontière linguistique (phrase, unité A ou clause) ; enfin, un autre quart des pauses très longues sont situées à l'intérieur d'une clause, sans qu'il y ait de révision. Les pauses à l'intérieur des clauses se situent plutôt en début de clause, par exemple entre la conjonction de subordination et le sujet de la clause (« tant qu' **(16'') il** parvenait à sortir d'ici »[77]) ou entre le sujet et le verbe (« Mais au moment où il **(18'') allait** s'endormir »). Certaines pauses très longues correspondent sans doute à des hésitations quant au bon lexème à utiliser, la solution choisie étant d'ailleurs parfois modifiée a posteriori. Par exemple, l'expression « le pauvre gardien ébahi » contient une pause de 15 secondes avant « le » et une nouvelle pause de 21 secondes avant « pauvre » ; hors l'adjectif « pauvre » est raturé par la suite, et l'on sait que ce scripteur porte une attention particulière au choix de la désignation des référents, puisqu'après la modification de la phrase précédente, cette expression sera remplacée par « ce dernier stupéfait ». De manière similaire, le prédicat « fut assuré d'être » qui est ensuite remplacé par « put » puis « fut certain d'être » est précédé d'une pause de 15 secondes, le scripteur hésitant sans doute sur la meilleure manière de formuler son prédicat, même si le contenu sémantique était déjà planifié.

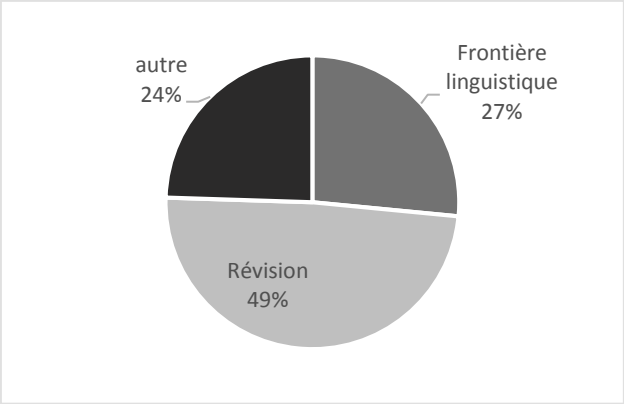


Figure 73. Contenu des pauses D10 (en %)

2.5.2. Exploration des questions soulevées

2.5.2.1. Reformulation du premier passage (Q1)

Pour les trois étapes de reformulation du premier passage, nous présenterons en détail sous forme de tableaux les durées des pauses durant les sessions de révision, et nous indiquerons entre parenthèses les pauses Q4 dans le contexte entourant la session de révision.

- Etape 1 :

[77] Dans les exemples de cette partie sur l'étude des pauses, seules les pauses Q4 (parmi les 25% de pauses les plus longues) sont indiquées entre parenthèse avant le mot qu'elles précèdent, indiqué en gras.

« D'un bond, il se jeta sur l'homme et lui arracha les clés que ce dernier tenait encore **(7,5'')** dans sa main. [rev] **(8,4'')** [RL] puis »

[rev] :

14030	ins mot	de
1112	ins mot	la
1060	sup ins constituant	dans sa
2228	sup constituant	que ce dernier tenait encore
3847	remp mot	de > des
556	remp mot (gram)	main > mains
1216	sup mot	la
600	esthé	.
3487	esthé	rature "la"
804	esthé	rature "que ce dernier tenait encore"
21985	remp pct	. > , (retour origine)

Tableau 69. Révisions en début du texte - Détails de l'étape 1 avec durée des pauses (en ms)

Nous avions vu que cette première étape concernait essentiellement le problème de la structure du prédicat (tenir dans sa main / tenir de la main / arracher des mains). Il est possible que cette hésitation ait commencé à apparaitre dès l'écriture de la phrase, puisqu'on relève une pause de 7,5 secondes avant l'écriture de « dans sa main ». Une longue pause précède la session de révision, mais les différentes modifications s'enchainent ensuite relativement vite (les pauses n'appartiennent pas à Q4, même si certaines durent près de 4 secondes). Il semble donc que ce soit en effectuant le premier remplacement que le scripteur ait l'idée de finalement supprimer un constituant, ce qui le conduit à modifier la structure du dernier groupe nominal. L'absence de pauses très longues durant ces opérations suggère que le scripteur parvient à réfléchir à ces modifications en parallèle de l'exécution. La deuxième partie de cette étape, consistant à modifier la ponctuation constituant la séquence où la révision est initiée, est également précédée d'une longue pause (22 secondes, auxquelles on peut ajouter les 5 secondes cumulées précédant les reprises esthétiques).

- Étape 2 :

« sans le remarquer, **(37'') hurlant** [rev] **(14'')** Il »

[rev] :

15830	esthé mot	disparut
2895	esthé mot	et
7456	aj mot	aussitôt
292	aj constituant	,
4216	sup mot	pauvre
24541	sup mot	hurlant
796	remp pct	, > . (retour origine)

Tableau 70. Révisions en début du texte - Détails de l'étape 2 avec durée des pauses (en ms)

Durant cette étape, il y a 2 ajouts et une suppression au début du texte, qui s'enchainent sans pauses très longues. En revanche, la révision a été très longue à débuter : les modifications sont précédées d'une pause de 7,5 secondes, qui intervient après deux reprises esthétiques. Or, la pause de 16 secondes avant la première reprise esthétique n'est sans doute pas due à la planification de celle-ci, mais est un indice de la relecture de ce début de texte et d'une hésitation sur ce passage. Les opérations d'ajout interviennent d'ailleurs juste après le mot « disparut », qui est le premier mot sur lequel le scripteur a laissé une trace. Par la suite, une longue pause précède les modifications apportées sur la phrase précédant la session de révision. Outre une possible relecture pour s'assurer de la cohérence apportée par les modifications, cette longue pause sert vraisemblablement à réfléchir aux deux modifications à apporter à l'emplacement de l'origine, puisque le remplacement de la ponctuation se fait très rapidement après la suppression du mot « hurlant ». La planification de la suite se fait sans doute après, lors de la pause de 14 secondes précédant l'écriture de la nouvelle phrase.

Il pouvait paraitre surprenant qu'une révision du début du texte soit initiée à cet endroit du texte, alors qu'il n'y avait pas de lien sémantique particulier. Là encore l'étude des pauses peut donner un indice : le participe présent « hurlant » est précédé d'une pause de 37 secondes, ce qui est très long. Cela suggère que ce participe présent n'était pas anticipé dès le début de la phrase, et que son écriture a fait l'objet d'une longue hésitation. Doutant de son choix, le scripteur a alors sans doute entrepris une relecture du début du texte pour savoir comment poursuivre au mieux son texte, et a alors perçu la nécessité d'apporter des petites corrections à ce passage. Suite à ces différentes relectures (sans doute avant l'écriture de « hurlant », avant les modifications puis avant la suppression de « hurlant »), le scripteur a décidé de modifier le contenu envisagé. En effet, aucun élément de la phrase qu'il écrit ensuite ne reprend cette idée de hurlement : il ne s'agit donc pas d'une simple reformulation, mais d'un changement au niveau de la macro-planification.

- Étape 3 :

« sa chambre, sa [rev] tranquillité »

[rev] :

37527	sup const	d'un bond, il se jeta sur l'homme et lui arracha les clés
684	sup mot	des
300	sup mot	mains
2603	sup mot	puis
11451	ref mot	ses > les
872	aj mot	du
1136	aj mot	gardien
3932	aj mot	ce
985	aj mot	dernier
492	aj pct	,

1176	aj mot	stupéfait
104	aj pct	,
3316	sup const	le gardien ébahi
832	esthé	ce
2220	esthé	stupéfait
27448	esthé	rature "puis"
2344	aj mot	Leste
1124	aj mot	comme
384	aj mot	un
1856	aj mot	chat
212	aj pct	,
17761	esthé mot	sa (retour origine)

Tableau 71. Révisions en début du texte - Détails de l'étape 3 avec durée des pauses (en ms)

L'analyse des pauses fait ressortir trois étapes au sein de cette session de révision. Tout d'abord, une très longue pause (près de 40 secondes) aboutit à la suppression de la première partie de la phrase, ainsi que du connecteur qui reliait cette première partie à la clause suivante. Une seconde pause longue précède l'ensemble des reformulations lexicales qui portent notamment sur le référent « le gardien », afin de permettre une continuité référentielle adéquate entre les propositions. Enfin, après une nouvelle longue pause, le scripteur ajoute le constituant qui sera le début de son texte (« Leste comme un chat, »). La reprise de l'écriture du texte est précédée d'une longue pause sans doute dévolue à sa planification, comme en témoignent les 18 secondes précédant la reprise esthétique du mot « sa », d'où la révision avait été initiée.

2.5.2.2. Évolution du scénario (Q2)

L'étude de l'ordre des révisions avait révélé que le remplacement de « sa maison » par « ses parents » précédait de très peu le groupe nominal « le petit Dimitri » : il serait donc possible que le changement de scénario soit imaginé à peu près à ce moment. Un indice supplémentaire peut être donné par la durée des pauses.

Le passage avec les pauses longues (Q4) est :

« Mais au moment où il **(18,2'') allait** s' (**7,4'') endormir**, le téléphone sonna [rev] **(36,3'') . (87,2'') «** Comment ? Le petit Dimitri ? Il a disparu ? **(13,5'') Mais** ses parents viennent le chercher demain matin ! »

[rev] :

5339	esthé	téléphone
23797	remp	sa > ses
2027	remp	maison > parents

Tableau 72. Révisions autour de la mention "petit Dimitri" - Détails avec durée des pauses (en ms)

Il apparait que ce passage comporte plusieurs pauses très longues, dont la plus longue du texte (84,2''). Tout d'abord, il est remarquable que la phrase précédant la révision compte deux pauses longues, voire très longues, qui en outre ne sont pas situées en début de phrase ou de clause. La première pause (18,2'') se situe avant le verbe, ce qui est le cas d'autres longues pauses dans le texte. Cependant, le fait que le scripteur s'arrête de nouveau plus de 7 secondes entre l'auxiliaire et l'infinitif « endormir » montre que la réflexion porte au-delà du simple prédicat de la clause. En effet, il serait étrange que le scripteur hésite 7 secondes à propos du verbe « endormir », alors que le pronom « s' » était déjà écrit. Au total, il y a plus d'une minute de pause entre la fin de « sonna » et le remplacement de « sa » par « ses ». Une réflexion intense est vraisemblablement engagée durant cette période, sans doute à propos de la pertinence de présenter un Dimitri enfant. Il y a ensuite 87,2 secondes de pauses avant le discours rapporté, alors que la décision de parler du « petit Dimitri » a déjà été prise. Il est possible que ce long temps permette d'affiner le nouveau scénario, notamment concernant le fait que Dimitri ne se trouve pas dans une cellule, mais dans un établissement médical. Bien que le contenu exact des réflexions menées pendant ces longues pauses ne nous soit pas directement accessible, il est fortement probable qu'elles portent sur le contenu du scénario modifié en cours d'écriture.

2.6. Discussion

2.6.1. Réponse aux questions soulevées

2.6.1.1. Révisions concernant la formulation (Q1)

Notre première question était : dans quelle mesure la correction syntaxique et orthographique de ce texte a-t-elle été retravaillée par le scripteur ?

La part de révisions consacrée aux corrections orthographiques est très faible dans ce texte, puisque seules 3 occurrences sont relevées :

- une effectuée à la fin de la phrase : « à pas de loups > à pas de loup »
- un oubli d'apostrophe (« s > s' ») corrigé dès le mot suivant, le scripteur interrompant l'écriture de celui-ci après la première lettre du mot
- une effectuée lors de la relecture finale (« papillottant > papillotant)

Comme aucune erreur d'orthographe n'a été relevée dans le produit final, cela suggère que ce scripteur a un très bon niveau d'orthographe, puisqu'il produit peu d'erreurs dans son jet initial, et les corrige rapidement. Cependant, durant la relecture finale une part de l'attention est tout de même portée sur l'orthographe, puisque l'erreur qui était encore dans le texte est corrigée, la détection de celle-ci pouvant être difficile car il s'agit d'un problème de double consonne dans un mot peu fréquent.

En revanche, ce scripteur n'a pas hésité à effectuer des modifications sur son texte. Celles-ci concernaient différents niveaux du texte : remplacement d'un terme par un synonyme, changement de

la structure de la phrase, ajouts et suppressions de constituants (groupes nominaux, phrases), changement de segmentation des unités. Il est à noter que les modifications nécessaires étaient apportées pour garder un texte cohérent, comme l'a montré le travail de reformulations lexicales du référent « le gardien » dans le premier passage. En outre, ces moments de révisions ont souvent été initiés longtemps après le passage modifié, sans doute à l'occasion d'une relecture. Dans plusieurs cas, le travail effectué sur un passage et la relecture sans doute pratiquée à cette occasion a conduit le scripteur à modifier au niveau de la macro-planification le texte qu'il était en train d'écrire.

En outre, concernant la gestion temporelle de ces processus, nous pouvons noter que de nombreuses révisions, voire sessions de révisions, sont initiées en milieu de clause. Les révisions de ce type relevées dans le corpus portaient généralement sur un mot très proche, soit pour en corriger l'orthographe, soit pour le reformuler. Dans le cas de ce texte, le scripteur s'engage régulièrement dans une révision beaucoup plus large, la séquence affectée étant parfois très loin de l'origine.

2.6.1.2. Révisions concernant le changement de scénario (Q2)

L'hypothèse d'un changement de scénario, proposée à partir du produit final, n'a pu être vérifiée que grâce à la complémentarité des indices à notre disposition. En effet, la seule étude du manuscrit n'infirmait pas cette hypothèse, mais elle nous a conduit sur des fausses pistes : par exemple, nous avions supposé que les changements liés à la première partie avaient été faits dans le but de rendre l'ensemble plus cohérent, alors que ce n'était pas le cas. Au contraire, la chronologie des révisions nous a plutôt conduit à l'hypothèse, celle-ci ne pouvant pas être vérifiée, que ce sont justement ces révisions du début du texte qui ont fait évoluer la représentation du personnage, et ont peu à peu conduit à imaginer un personnage principal qui était en réalité un enfant. En examinant la localisation de l'origine des révisions et la durée des pauses entourant certaines opérations, nous sommes parvenus à déterminer l'endroit où, vraisemblablement, une grande partie du scénario a été replanifiée. Nous sommes ici plutôt dans une écriture « romantique » ou « à processus », durant laquelle le scripteur n'hésite pas à retravailler son texte au fur et à mesure de la production, de nouvelles idées émergeant au cours de l'activité de rédaction.

2.6.2. Corrections et reprises esthétiques : des indices de relecture ?

Les analyses menées précédemment sur certains passages ont montré qu'à plusieurs reprises, les modifications apportées sur le texte ont aidé à planifier la suite du texte qui était en train d'être écrit, par exemple en conduisant à une modification de la ponctuation ou à la suppression du mot qui avait été écrit avant la séquence de révision. Il semble que la lecture du texte déjà écrit est, du moins pour ce scripteur, un outil pour la planification de la suite du texte. En l'absence de l'enregistrement des mouvements oculaires, il n'est pas possible de retracer l'ensemble du parcours de lecture du scripteur.

Cependant, un texte comme celui-ci comportant de très nombreuses séquences de révisions peut déjà renseigner sur ce qui a été relu. En effet, on peut partir du principe qu'une trace laissée sur un mot indique que, au minimum, le regard s'est porté sur ce mot. Ainsi, l'étude des reprises esthétiques peut notamment renseigner sur les endroits où le scripteur est « repassé », et le moment où cette relecture a été initiée.

En tout, 43 reprises esthétiques ont été relevées. Dans plus de la moitié des cas (26/43), la reprise esthétique suit immédiatement l'origine de la révision, c'est-à-dire que c'est la première – voire la seule – opération de révision effectuée. Parfois, ces reprises esthétiques s'inscrivent dans une session de révision : elles peuvent précéder une reformulation, comme avant la modification « sa maison > ses parents », soit précéder la reprise de l'écriture après une longue session de révision, comme lors du changement de ponctuation après la première session de révisions portant sur les deux premières lignes du texte. Dans certains cas il n'y a aucune correction avant ou après cette reprise esthétique : cette relecture peut alors simplement être une aide pour écrire la suite du texte. Ainsi, dans l'exemple (145), des longues pauses débutent la deuxième phrase : 7,1'' avant le sujet, puis 27,2'' avant le verbe « voyait ». Durant cette deuxième période de pause, il est fort probable que la phrase précédente ait été relue, puisque l'on sait au moins que le scripteur a posé son stylo sur le premier mot (« Dimitri ») de cette phrase.

(145) ***(6,5'')*** **Dimitri** ***(8,32'')*** **avai***t bon espoir qu'il s'endorme, il attendrait* ***(7,9'')*** **aussi** longtemps que nécessaire. (***7.1'') Il (15,9'') [esth «*** **Dimitri** ***»] (11.3'')*** voyait *l'homme lutter* contre le sommeil,

L'observation de la distribution spatiale des reprises esthétiques est très informative pour ce texte, que ce soit pour expliquer l'origine de la reprise esthétique ou son site. La Figure 74 présente uniquement les reprises esthétiques effectuées en interrompant le texte en cours d'écriture, qu'elles soient ou non suivies de modifications. Ces révisions sont initiées à 10 reprises en fin de ligne. Dans la majorité de ces cas il s'agit également d'une frontière unité linguistique, mais il arrive que cette reprise interrompe la clause, par exemple après la préposition « derrière ». Dans le texte, il arrive à 6 autres occasions qu'une reprise esthétique soit initiée après une ponctuation. Dans seulement deux cas la reprise esthétique est située « loin » de son origine : entre « hurlant » et « disparut », qui sera le début d'une des sessions de révision du début du texte, et entre « Dimitri » et « peu importaient ». L'affirmation que la plupart de ces reprises esthétiques sont « proches » de l'origine diffère du résultat mentionné précédemment, qui indiquait que, contrairement au corpus dans lequel la plupart des reprises esthétiques se situaient sur le mot précédent, et généralement à une frontière linguistique, les reprises esthétiques de ce texte portaient souvent sur un endroit plus éloigné dans le texte (distance « courte » ou « moyenne »). Par exemple, la plupart des reprises esthétiques initiées en fin de ligne portent sur le premier mot de cette ligne. Or, si le scripteur a déjà amené son stylo à gauche de la page pour écrire la ligne suivante,

le premier mot de la ligne est en réalité le mot le plus proche ; d'ailleurs, il y a un cas où, après avoir écrit la conjonction « et » en début de ligne, le scripteur fait une nouvelle reprise esthétique sur le mot « certain », premier mot de la ligne supérieure. De manière générale, on voit que les reprises portent sur quelques mots « avant » la séquence d'origine, avec une relecture qui se fait de manière horizontale. Cependant, il peut arriver que le mot le plus « proche » le soit verticalement : c'est ce qui peut facilement expliquer la reprise esthétique sur « tranquillité » lorsque le scripteur s'arrête après « couloir », alors qu'il serait plus difficile de fournir une explication d'ordre linguistique du rapport entre ces deux mots. Les reprises esthétiques étant donc situées spatialement proches du point d'origine, mais étant en général très longues (54% sont des pauses Q4), elles sont l'indice d'une relecture qui s'étend vraisemblablement bien au-delà de la séquence révisée[78].

Un des outils qui n'était malheureusement pas à notre disposition pour ce texte, mais qui aurait pu fournir des informations supplémentaires, est donc l'enregistrement des mouvements oculaires. La Figure 75 donne un exemple d'un « aperçu des mouvements oculaires » d'un texte issu d'une autre collecte de données. Le texte du scripteur S31 ne permet pas de donner autant de détails, mais certains « chemins de relecture » peuvent être reconstruits à partir de la localisation des origines et sites des révisions, comme illustré par la Figure 76. Dans les deux cas, nous voyons des traces de mouvement (oculaire ou de stylo) « verticales », c'est-à-dire des cas où le scripteur relit le texte déjà écrit plusieurs lignes auparavant.

[78] Le terme « reprise esthétique », choisi au début de nos analyses, partait du postulat que ces révisions avaient pour principal but d'améliorer la lisibilité du texte. Les analyses présentées ici conduisent davantage à les considérer comme des marques laissées lors de relectures ; dès lors, un terme plus adéquat que « reprise esthétique » serait à envisager.

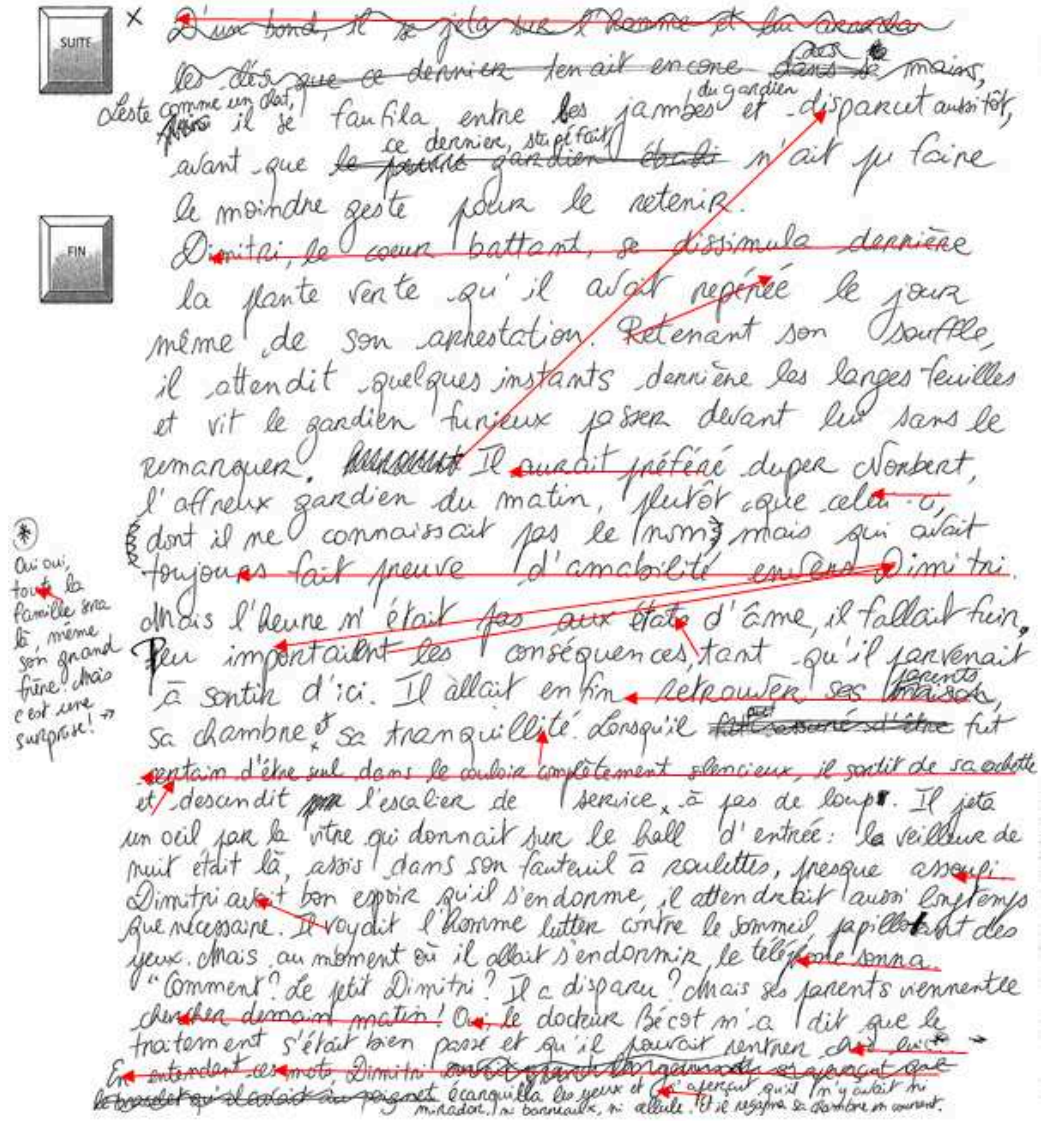


Figure 74. Reprises esthétiques initiées durant l'écriture du texte du scripteur S31[79]

[79] Le début de la flèche indique l'origine de la révision ; la pointe de la flèche indique le site de la révision.

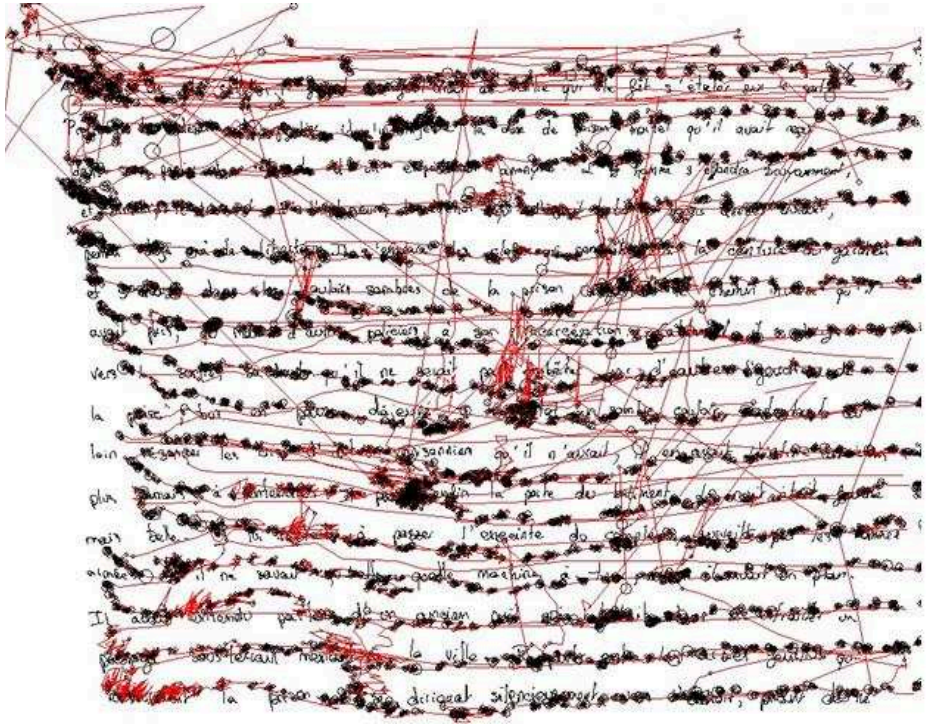


Figure 75. Capture d'écran du logiciel Eye & Pen - Aperçu des mouvements oculaires recueillis lors de la production d'un texte

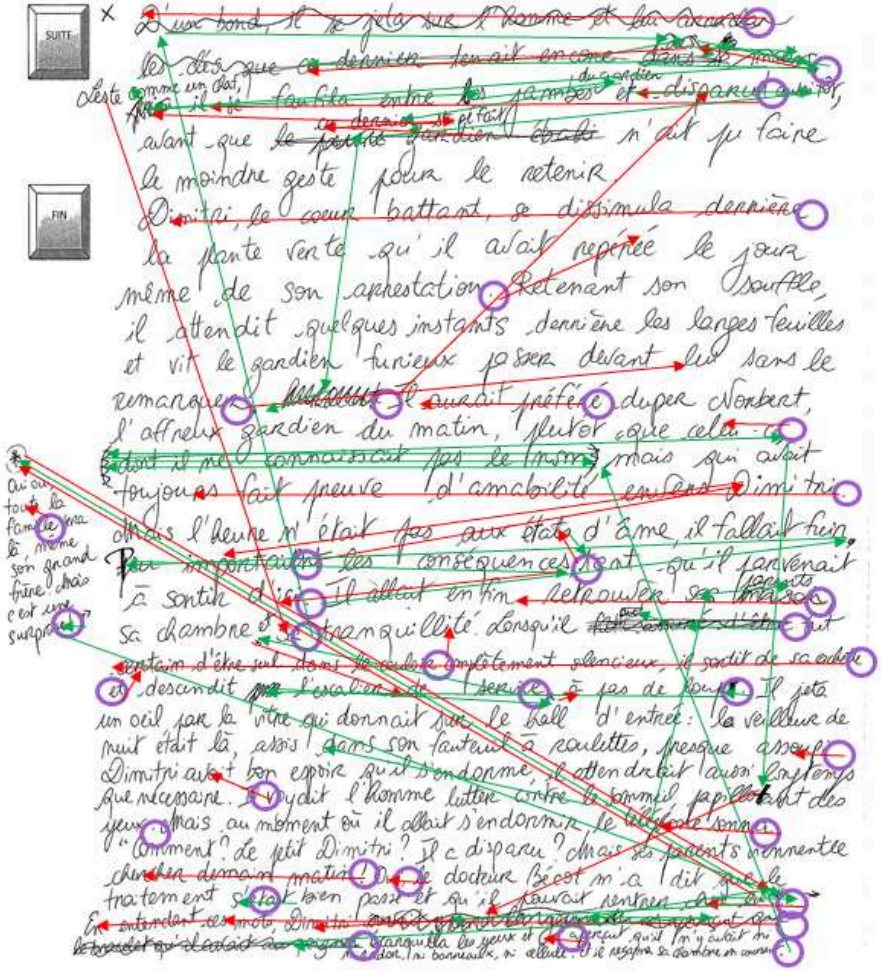


Figure 76. Origine et site des révisions du texte du scripteur S31[80]

[80] Les cercles violets indiquent les origines des révisions interrompant l'écriture du texte ; les flèches rouges représentent les reprises esthétiques, et les flèches vertes les révisions avec modification du texte déjà écrit.

2.6.3. Les outils à disposition

L'étude de cas telle que nous l'avons présentée nous a permis de mettre en évidence, étape après étape, quelles étaient les traces à notre disposition en fonction du matériel utilisé, et surtout de quels processus ces empreintes et indices étaient les témoins. Pour rappel, dans le cadre de notre étude, nous utilisons le terme de « trace » pour désigner « un objet inscrit dans une matérialité que nous percevons dans notre environnement extérieur et dotons d'un potentiel de sens particulier » (Jeanneret, 2010, p. 61). Si cela correspond à « la capacité dans le présent de faire référence à un passé absent mais postulé », nous parlons d'« empreinte » ; si la trace renvoie à « des choses qui arrivent en même temps sans être pour autant visibles en même temps » (Krämer, 2010, p. 14), nous parlons d'« indice ».

Au niveau du manuscrit, le texte écrit était l'empreinte de choix lexicaux et syntaxiques, et était l'indice d'une macro-planification (e.g. caractéristiques des personnages, du cadre, évolution possible de ce scénario au cours de l'écriture du texte). Les modifications apportées au texte déjà écrit (e.g. ratures, ajouts) étaient l'empreinte d'un changement de choix lexicaux ou syntaxiques, et étaient donc l'indice d'une réflexion (i.e. micro-planification ou formulation). Les marques de ponctuation étaient, selon nos hypothèses, des indices d'unité de planifications.

L'écriture enregistrée a permis tout d'abord d'avoir à notre disposition l'ensemble des révisions, ainsi que la temporalité liée à celles-ci (ordre des révisions, lieu d'origine de la révision, durée). Là encore, les modifications apportées au texte déjà écrit, dont le relevé était plus complet que celui obtenu à partir du manuscrit, étaient l'empreinte des changements de choix lexicaux ou syntaxiques, et donc l'indice de réflexions. Plus particulièrement, la temporalité relative (i.e. l'ordre des révisions) donnait des indices quant aux raisons de certains choix : par exemple, la nécessité de changer d'items lexicaux pour exprimer de manière plus adéquate la continuité référentielle modifiée suite à d'autres changements. Les lieux d'origines des révisions (i.e. l'endroit où la révision a été initiée) donnaient des indices quant à la gestion des multiples processus au sein des unités de production : il est ainsi possible de repérer si le scripteur interrompt sa production pour effectuer des modifications au niveau des choix lexicaux ou syntaxique ; s'il s'interrompt pour corriger des erreurs d'orthographe repérées ; s'il effectue des retours systématiques sur le texte, soit en frontière linguistique, soit pour des raisons d'organisation spatiale (e.g. en fin de ligne). Enfin, les « reprises esthétiques » sont l'empreinte d'un retour sur le texte ; associées à la durée des pauses et aux lieux d'origine des révisions, elles deviennent l'indice d'une relecture du texte déjà écrit. Par ailleurs, les durées des pauses, empreintes d'un arrêt plus ou moins long, peuvent être associées à d'autres éléments pour devenir des indices : associées aux emplacements syntaxiques, elles sont des indices des unités de production ; associées au type de révision effectué, elles sont des indices de traitement cognitifs plus ou moins couteux, parfois non visibles par le simple examen des révisions (e.g. pause très longue étant un indice de macro-planification, pause longue avant une reprise esthétique étant un indice de relecture plus importante).

Malheureusement, nous n'avons pas à notre disposition l'enregistrement des mouvements oculaires, qui auraient permis de compléter nos analyses. En effet, les mouvements oculaires portés sur le texte déjà écrit sont des empreintes de relecture. Associés au contexte (lieu où le retour en arrière est initié, segment relu, segment écrit par la suite), ces mouvements oculaires seraient des indices des processus de réflexion, soit en vue de planifier ce qui va suivre, soit en vue de modifier ce qui a été déjà écrit.

3. Discussion générale

L'étude des révisions apporte donc un nouvel éclairage sur les résultats et hypothèses présentés dans les chapitres précédents, et ouvre de nouvelles perspectives de recherche.

Tout d'abord, l'étude de cas rappelle l'importance de prendre en compte la variabilité interindividuelle, certes en présentant un cas assez extrême. Cependant, le fait d'avoir un texte comportant 90 séquences de révisions, alors que les textes produits avec les mêmes consignes, et de longueur équivalente, en comptent entre 1 et 15, montre qu'il est difficile de généraliser les résultats présentés pour un type de scripteurs. De même, la mise en lien entre les données chronométriques et linguistiques avaient permis de mieux comprendre le décours temporel des processus de micro-planification et de formulation, mais il ne faut pas oublier qu'il s'agissait de phrases ne comportant pas de révision ; un tel type d'étude n'aurait pas été pertinent pour le texte du scripteur S31.

Les révisions du corpus suggèrent que, dans le cas d'un texte peu modifié, les modifications se font plutôt au cours de l'écriture de la clause, et sur un petit empan. À l'inverse, le scripteur S31 n'hésite pas à faire de longues sessions de révisions, celles-ci se situant parfois loin du point d'origine, même si ce point d'origine se situe au milieu d'un syntagme. La localisation des origines des révisions indique tout de même une préférence pour les frontières d'unité linguistique. Dans le corpus, ces pauses en frontière témoignent d'ailleurs des difficultés éprouvées par le scripteur pour marquer de manière correcte la frontière de l'unité A. En outre, même s'il n'est pas certain que les scripteurs aient systématiquement relu leur clause dans un but de vérification à la fin de l'écriture de celle-ci, il est possible que ces frontières de clause soient l'occasion d'une relecture du texte déjà écrit, comme en témoignent les nombreuses reprises esthétiques, qui sont précédées d'une pause longue tout en étant spatialement proche.

Ces « reprises esthétiques », peu mentionnées dans la littérature, sont en effet le témoin d'un passage sur le texte, et la longueur qui leur est associée suggère que le traitement est assez couteux ; or, elles n'aboutissent pas une modification de la séquence. Leur analyse ne peut se faire sans se référer à la segmentation linguistique et à la disposition spatiale du texte : généralement, elles sont initiées en fin de ligne ou à une frontière linguistique ; la séquence « affectée » par cette reprise esthétique est très souvent proche, même s'il ne s'agit pas nécessairement du mot précédent (e.g. mot au début de la ligne dans le cas d'un retour à la ligne ; mot située sur la ligne supérieure).

Une des perspectives soulevées par l'étude de toutes ces révisions concerne justement le rôle de la relecture lors de la production. La frontière est en effet parfois ténue entre « relecture en vue d'avoir des idées pour poursuivre le texte » et « relecture en vue de vérifier l'adéquation entre ce qui est écrit et la représentation mentale ». Ainsi, il est arrivé que des relectures aboutissant à des modifications du texte déjà écrit conduisent également à une modification de la dernière séquence produite avant le retour sur le texte, comme si la relecture avait fait évoluer ce qui avait été planifié. Une autre illustration assez

forte de ce lien entre relecture et planification vient également de l'étude de cas : la longue session de révisions ayant abouti aux modifications importantes du début du texte a été initiée après un déterminant, au cours d'une énumération. Il était peu probable que le scripteur envisage alors une relecture en vue d'une vérification ; il souhaitait peut-être, en revanche, s'aider de ce qui avait été déjà écrit pour trouver le terme manquant. Or, la relecture a conduit à modifier le début du texte ; le texte en cours d'écriture n'a pas été modifié, mais il est possible que les changements apportés sur la phrase initiale aient changé la représentation mentale du personnage, contribuant ainsi au changement situé au niveau de la macro-planification observé dans la suite du texte. Bien sûr, cette perspective concernant la relecture gagnerait à être explorée par l'utilisation de l'enregistrement des mouvements oculaires.

Conclusion

1. Synthèse des résultats

1.1. Les unités de production : de la clause à la phrase

1.1.1. Aspect développemental : vers la production d'unités supra-clausales

Un des buts de notre recherche était d'identifier un pattern développemental concernant la taille de l'unité de production. Conformément à des études antérieures, il est apparu que les plus jeunes (élèves de CM2) planifiaient plutôt clause par clause, faisant pour la plupart des pauses très longues majoritairement en frontière de clause, même lorsque celle-ci était interne à l'unité T. En revanche, cette unité T semble devenir une unité de production davantage pertinente pour les élèves de 5ème, et surtout de 3ème. L'allongement de ces pauses à des emplacements syntaxiques élevés dans la hiérarchie (débuts d'unités T ou de phrases) est certainement l'indice d'une activité cognitive plus couteuse. Par ailleurs, nous avons relevé que les scripteurs maitrisant bien les normes de segmentation font davantage leurs longues pauses en frontière d'unité T et de phrase, alors que le fait de faire des longues pauses en frontière de clause est souvent associé à un manque d'application des normes typographiques de segmentation : ainsi, il semble bien qu'il y ait un lien entre maitrise de la segmentation et capacité à envisager une planification supra-clausale.

Les résultats obtenus concernant l'impact de l'ordre de production et du type de texte sur la gestion temporelle de l'écriture témoignent de l'intérêt d'étudier le développement tardif du langage, plus particulièrement entre 10 et 15 ans. Nous avons relevé une augmentation de la quantité de texte produit, ainsi qu'une complexification de l'empaquetage syntaxique, avec une augmentation de la proportion de clauses subordonnées. Par ailleurs, le nombre d'unités T se rapproche du nombre de phrases, ce qui peut être mis en lien avec la meilleure maitrise des normes typographiques, puisqu'à l'issue du collège peu de participants présentent encore des problèmes liés à la ponctuation. Les résultats montrent une automatisation des processus de « bas-niveau », avec un raccourcissement des temps d'exécution et des pauses à l'intérieur des clauses, les élèves de CM2 étant certainement encore fortement contraints par ces processus. Les élèves de 3ème, quant à eux, mettent en place des stratégies parfois plus couteuses, mais qui permettent d'élaborer des textes correspondant davantage aux attentes. Finalement, il semble que les élèves de 5ème bénéficient le plus de l'ordre de production et du type de texte : il s'agit sans doute d'une période où les processus de bas-niveau sont assez automatisés, mais où l'élaboration concernant les processus de haut niveau est moins complexe, permettant ainsi aux scripteurs de consommer moins de ressources cognitives lorsque le texte a déjà été produit à l'oral, ou lorsque le type de texte est bien connu.

1.1.2. L'anticipation lors de la production de clauses syntaxiquement liées

L'effet syntaxique observé dans plusieurs études précédentes (e.g. Foulin, 1998; Schilperoord, 1996) a été corroboré par nos analyses. Plus précisément, nous avons pu établir une hiérarchie quant à la durée moyenne des pauses entre la clause, l'unité Autonome et la phrase. Les analyses prenant en compte les caractéristiques syntaxiques des liens entre les clauses ont suggéré que les clauses subordonnées étaient planifiées avant leur pause initiale, donc vraisemblablement pendant l'écriture de la clause matrice, ou pendant la pause initiale de celle-ci. Cette hypothèse a été corroborée par l'observation d'une anticipation de nombreux éléments linguistiques de la clause subordonnée, notamment syntaxiques, pendant l'écriture de la clause précédente (clause matrice). Ce qui est apparu plus surprenant est que, peut-être justement à cause de cette anticipation de la clause subordonnée, la variance des données chronométriques de la clause précédant la clause matrice (appartenant donc à une autre unité T) était fortement expliquée par des éléments linguistiques de la clause matrice. La présence de liens syntaxiques forts amène vraisemblablement le scripteur à anticiper le contenu linguistique des clauses suivantes. Par ailleurs, il a été observé que la variation du temps d'écriture des clauses autonomes (i.e. ni clause matrice ni subordonnée) était peu expliquée par les caractéristiques linguistiques de cette clause. Si le cout cognitif de la production de ces clauses n'est pas lié à la micro-planification et à la formulation de la clause autonome, il conviendra de s'interroger sur les traitements effectués en parallèle de la production de ce type de clauses, et qui pourraient davantage expliquer ces variations. Une des hypothèses serait que l'absence de lien syntaxique fort avec les autres clauses facilite la production, libérant des ressources pour par exemple anticiper les clauses suivantes ou effectuer des relectures en parallèle de l'écriture.

Par ailleurs, nous attendions que les analyses des révisions, notamment du moment où elles étaient initiées, donnent des indices supplémentaires quant à cette unité de production : par exemple, que le scripteur relise de façon assez systématique l'unité produite, donc en fin de clause ou de phrase. Dans l'ensemble du corpus, nous avons tout d'abord constaté que peu d'erreurs d'orthographes étaient corrigées, et que ces corrections n'étaient pas majoritairement effectuées à partir d'une frontière d'unité linguistique. Cependant, au niveau des frontières de phrase, nous avons relevé de nombreuses « reprises esthétiques », que nous avons par la suite interprétées comme des indices de relecture, suggérant donc des retours sur le texte. La question reste ouverte concernant le but de ces relectures : la vérification « de surface » ne semblant pas très fréquente ou du moins pas très efficace, puisque de nombreuses erreurs d'orthographe persistent, on peut suggérer que ces relectures servent soit à une vérification « en profondeur », notamment pour s'assurer de la cohérence de l'unité produite, soit à s'aider du texte déjà écrit pour produire la suite.

1.1.3. La phrase : unité de planification ou de production ?

Le fait de considérer la phrase comme une « unité de planification » tient essentiellement au fait que cette unité est précédée d'une longue pause chez les scripteurs adultes, ce qui est censé refléter une planification couteuse, donc portant sur un empan assez large. Or, si la longueur des pauses initiales de phrase est de manière indéniable longue – voire très longue – plusieurs résultats de nos analyses poussent à rester prudent quant à une interprétation en termes de planification de la phrase. Tout d'abord, l'observation des profils développementaux indiquait que le fait de faire des longues pauses en frontière de phrase était lié différemment à la maitrise des normes de segmentation : les élèves de 3ème faisant des longues pauses interphrastiques avaient peu de problème de ponctuation, affichant donc un profil similaire à celui des adultes, alors que les élèves de CM2 faisant des longues pauses interphrastiques avaient généralement des problèmes soit avec les majuscules, soit avec les virgules ; les processus engagés durant cette pause ne sont donc certainement pas les mêmes pour les deux profils de scripteurs. Par ailleurs, nous avons noté que la durée de la pause initiale n'était pas affectée par le nombre de clauses ou d'unités A dans la phrase ; de plus, la durée de production de la clause précédente, de la première clause de la phrase et la durée de la pause initiale de la phrase étaient peu affectées par les variables linguistiques caractérisant la phrase. Il est donc davantage probable que l'ensemble de la phrase ne soit pas planifié dès le début de cette phrase ; seuls quelques éléments sémantiques, notamment liés au prédicat, seraient planifiés avant le début d'écriture de la phrase. Les analyses qui avaient porté sur les clauses matrices et subordonnées suggèrent plutôt que la micro-planification et la formulation s'effectuent au niveau de l'unité T, et non de la phrase. La durée de la pause initiale pourrait alors être davantage due à des traitements portant sur l'unité précédente, comme la relecture ou l'intégration en mémoire à long-terme. Cela serait cohérent avec les « reprises esthétiques », généralement précédées d'une pause longue, observées majoritairement en frontière de phrase. La phrase ne serait donc pas une unité de « planification », mais pourrait tout de même être considérée chez les adultes comme une unité de « production », dans le sens où des éléments de planification plus généraux et des traitements liés à l'unité venant d'être produite se situeraient justement en frontière de phrase. Bien que l'ensemble de la phrase ne soit donc pas planifiée en détail dès le début de son écriture, le scripteur, en choisissant une ponctuation forte, marquerait bien la fin d'une unité considérée comme un « tout ».

1.2. La gestion des unités de production

1.2.1. La difficulté de segmenter le texte

Nous avions exposé, d'un point de vue théorique, les difficultés à définir ce qu'était une « phrase ». Nous avions finalement retenu des critères typographiques pour, en tant que lecteur, identifier ce que le scripteur avait voulu délimiter comme étant une phrase. Cependant, cette difficulté de segmentation

reste présente pour le scripteur. Nous avions choisi de considérer le système de ponctuation comme un système qui, au même titre que le système des connecteurs, servait à indiquer le degré de connexion entre deux clauses. Ceci s'est avéré pertinent car les difficultés concernent généralement l'un de ces deux éléments. Ainsi, nous avons observé que, même chez les élèves de 3ème, certaines unités T n'étaient reliées ni par une virgule, ni par un connecteur, un seul de ces éléments étant suffisant pour marquer correctement la connexion. Cela montre la difficulté, jusqu'à tard, d'intégrer la nécessité de marquer explicitement la connexion entre unités T. Quelques cas d'hésitation ont en outre été relevés dans le corpus « Dimitri » : la modification du connecteur (« et » ou « mais ») ou l'ajout d'une virgule au niveau de la frontière entre unités A (qui était l'unité syntaxique intermédiaire entre la clause et la phrase dans le corpus adulte). Au niveau de la segmentation en phrases, nous avons noté quelques cas où des scripteurs supprimaient des connecteurs (e.g. « et » ou « ou ») pour les remplacer par des points, ou vice versa. Par ailleurs, la longueur de certaines pauses en frontière d'unité A, marquée ou non par un point, peut justement s'expliquer par cette hésitation à clôturer la phrase. Dans la plupart des cas, ces hésitations quant à la marque de connexion entre unités Apparaissaient dès le début de la clause suivante, et étaient rapidement résolues : les adultes connaissaient donc vraisemblablement bien les règles liées aux marques typographiques, et ajustaient leur production pour, par exemple, mieux s'adapter au récepteur.

1.2.2. Quelques facteurs rendant la production plus couteuse

L'explication de la variance des données chronométriques par le contenu linguistique des clauses a permis de faire ressortir ce qui rendait une production plus couteuse cognitivement, ce qui transparaissait par un allongement des durées des pauses et un ralentissement de la vitesse d'écriture. Plusieurs de ces éléments sont d'ordre discursif ; cela est important à souligner, car ils n'auraient sans doute pas été mis en évidence par la production d'une phrase isolée. En outre, ils témoignent de la contrainte pesant sur le scripteur pour produire un texte qui soit cohérent. Ainsi, une des raisons du ralentissement de la production est le fait d'évoquer un élément saillant du texte : présence du personnage principal, prédicat de 1er plan ; ces clauses, vraisemblablement essentielles pour la cohérence du récit, demandent un plus grand effort de micro-planification. Par ailleurs, le fait que des unités soient fortement liées entre elles, notamment au niveau sémantique, conduit généralement à un allongement du temps de production : c'est le cas lorsque l'indice d'analyse sémantique latente (i.e. similitude sémantique) est élevé, lorsqu'il y a un adverbe de liaison ou lorsque le sujet est maintenu d'une clause à l'autre. Le lien sémantique ne crée par une amorce facilitatrice, comme ce à quoi nous aurions pu nous attendre, mais au contraire augmente le cout puisque le scripteur est contraint de s'assurer que la cohésion est correctement établie (e.g. choix des bons items lexicaux pour assurer la continuité référentielle, absence de contradiction, respect des liens logiques). Nous avons en outre noté une production plus lente lorsque le prédicat était complexe, c'est-à-dire formé à l'aide d'un verbe modal ; le fait que le prédicat soit fréquent dans le

corpus facilite au contraire sa production, sans doute parce qu'il est davantage prévisible. Enfin, les compléments indiquant le cadre sont gérés différemment : le fait d'insérer un complément de temps « balise » accélère la production, alors que celle-ci est ralentie par la présence d'un complément de lieu.

1.2.3. L'interruption de l'écriture de la clause

La façon dont le scripteur gère l'écriture de sa clause peut en outre être mieux comprise grâce à l'étude des révisions : celles-ci sont des traces directes d'hésitations. Elles peuvent donc indiquer par exemple à quel moment il y a des incertitudes quant au choix d'un mot ou de l'orthographe. Nous avons relevé tout d'abord des retours assez « systématiques » dans le texte au fur et à mesure de l'écriture de la clause, permettant par exemple de corriger les erreurs portant sur des mots qui viennent d'être écrits (dans 75% des cas, les corrections d'erreurs sont effectuées au cours de l'écriture de la clause) ; les scripteurs sont même assez souvent prêts à interrompre le mot en cours d'écriture pour corriger l'erreur repérée. Il est cependant possible que ces détections et corrections d'erreurs ne soient pas dues à des traitements « automatiques », mais fassent plutôt suite à des relectures plus importantes de ce qui a été écrit auparavant : ceci serait corroboré par la longueur des pauses précédant ces corrections.

Concernant l'aspect incrémental de la formulation, nous avons relevé que, même dans le texte comportant de très nombreuses révisions, la clause est globalement bien planifiée, seules des petites modifications (inversions de lettres, choix d'un synonyme) étant généralement apportées, et ce avant que le mot ne soit complètement écrit. Bien sûr, le fait d'écrire sur papier, avec des ratures visibles, peut contraindre le scripteur à éviter d'apporter des changements importants par rapport à ce qui était écrit ; il serait alors intéressant de comparer cet aspect de la gestion d'écriture avec ce qui se passe lors de l'écriture par ordinateur.

1.3. Les données temporelles comme indices des processus rédactionnels

1.3.1. La prise en compte de différentes données temporelles

Nous avons souhaité utiliser les données temporelles à notre disposition, mises en lien avec le texte écrit, comme indices des processus rédactionnels. Parmi les « données temporelles », nous avons tout d'abord pris en compte les données chronométriques. Pour chaque mot, nous avons distingué les pauses initiales, les pauses intra-mots et l'exécution motrice ; cela nous a permis d'avoir pour chaque unité (e.g. phrase, clause) la pause initiale de l'unité, les moyennes des pauses inter-mots, des pauses intra-mots et du temps d'exécution. Mis en relation d'une part avec l'emplacement syntaxique (i.e. début de phrase, d'unité T ou unité A, de clause), et d'autre part avec le contenu linguistique des unités, ces données

chronométriques ont permis d'identifier des unités de production, et de mieux saisir à quel moment certains choix sémantiques, syntaxiques ou lexicaux étaient effectués.

Par ailleurs, nous avons utilisé des « données temporelles » liées aux révisions. Ainsi, la temporalité relative des révisions a permis de comprendre certains choix de modifications linguistiques. L'emplacement des origines des révisions a donné des indications sur la gestion de certains processus rédactionnels (e.g. hésitation quant à la formulation, correction orthographique) au cours de la production des clauses. Enfin, la durée élevée des pauses précédant certaines révisions, qui ne serait pas explicable uniquement par la modification apportée (e.g. dans le cas des reprises esthétiques), constitue des indices de l'engagement du scripteur dans des traitements cognitifs couteux, tels que la relecture.

1.3.2. La variabilité interindividuelle

Une des difficultés concernant la recherche sur la production écrite est la prise en compte de la variabilité interindividuelle. Celle-ci peut être visible à travers des caractéristiques identifiées des scripteurs, comme leur âge. La prise en compte de profils de scripteurs concernant leur maitrise des normes de segmentation a permis d'affiner les descriptions faites : certes, certains profils étaient davantage associés à un niveau scolaire, mais nous avons relevé des différences au sein de ces niveaux. Par ailleurs, ces profils ont montré que, par exemple, faire des pauses longues entre les phrases ne relevait sans doute pas des mêmes traitements pour les scripteurs faisant peu d'erreurs de ponctuation que pour ceux en faisant beaucoup. La variabilité est également présente dans les corpus produits par des adultes. Par exemple, nous avons vu émerger deux profils concernant la proportion de temps dévolue à l'exécution : des scripteurs pour qui l'exécution motrice constitue 20% du temps d'écriture, et ceux pour qui cela constitue 10% du temps d'écriture. Ce type de variabilité est à prendre en compte lors des analyses, ce qui a pu être fait grâce aux modèles mixtes utilisés. Enfin, des styles ont été identifiés dans les études précédentes, que ce soit à partir de manuscrits d'écrivains, de brouillons d'écoliers ou de textes écrits en situation plus expérimentale : il y a un continuum allant de scripteurs planifiant essentiellement en avance, puis effectuant peu de révisions lors de l'écriture, jusqu'à des scripteurs laissant émerger de nouvelles idées au fur et à mesure de la production de leur texte, n'hésitant pas à modifier fréquemment ce qui a été déjà écrit. La méthodologie utilisée dans notre recherche (i.e. la mise en relation des données chronométriques et du contenu linguistique) nécessitait de travailler sur des séquences de texte sans révision ; cela a été possible car les textes de notre corpus « Dimitri » en comportaient peu. Cependant, l'étude de cas présentée dans la dernière partie de notre travail montre justement un style d'écriture très différent, ce texte ayant pourtant été produit dans les mêmes conditions que les autres. Il convient alors de considérer ces différences comme une richesse : les révisions relevées ont ainsi permis d'apporter d'autres indices relatifs à la gestion temporelle de la production écrite ; en outre, cela incite à toujours rester prudent quant à la généralisation de certains résultats.

2. Perspectives

Ce travail de thèse s'inscrivait dans la lignée des études utilisant les données temporelles pour mieux comprendre la gestion des processus impliqués dans la production de textes. Il a permis de mieux saisir quelles étaient les unités de production et comment les processus tels que la micro-planification, la formulation ou la révision étaient gérés au sein de ces unités, mais de nombreuses perspectives de recherches peuvent être envisagées pour approfondir différents points observés.

Tout d'abord, l'intérêt d'analyser des textes produits par des enfants et adolescents ayant entre 10 et 15 ans a permis de caractériser la manière dont se développaient les compétences scripturales du point de vue des unités de production. Cet aspect développemental mériterait d'être approfondi ; cela nécessiterait la collecte d'un nouveau corpus, grâce auquel des éléments ayant un effet sur l'expertise rédactionnelle, comme l'empan de travail ou la fluence d'écriture, pourraient être pris en compte. Par ailleurs, la mise en relation de la taille des unités de production et de la maitrise des normes de segmentation a permis de mieux prendre en compte les différences entre les scripteurs et de donner une source d'explication du développement observé durant la scolarité au collège. Cet aspect pourrait être approfondi en essayant de saisir au plus près les connaissances des scripteurs quant à l'usage de la ponctuation, par le biais de l'énonciation de règles ou par des exercices, afin de faire un meilleur lien entre connaissances liées à la segmentation et production. Une comparaison entre la segmentation effectuée au cours de la production et celle présentée lors de la version finale au lecteur, en passant par l'étude des révisions portant sur cet aspect, pourrait par ailleurs mettre en évidence les hésitations des scripteurs, et permettrait de distinguer les marques de segmentation servant à planifier de celles ajoutées pour le lecteur. Au niveau didactique, il serait intéressant de voir si le fait de présenter le système de ponctuation comme un outil de connexion, plutôt que d'insister sur la segmentation, aurait pour effet un meilleur respect des normes et une diversification dans l'usage des marques de ponctuation.

Par ailleurs, le fait d'étudier des clauses au sein d'un texte a permis de faire apparaitre des effets qui ne se seraient certainement pas manifestés lors de la production de phrases : par exemple, l'anticipation de l'écriture de certaines clauses, ou la relecture d'unités précédentes. En outre, cela a permis, lors de la description linguistique, de relever des éléments constitutifs de la production d'un texte, tels que le liage syntaxique et sémantique, et de faire ressortir les effets liés à ces contraintes. Une des difficultés de notre travail était d'ordre méthodologique, puisque les données étaient complexes et produites par des scripteurs différents, sans contrôle de ce qui était produit. Les analyses utilisées ont cependant permis de prendre en compte la variabilité interindividuelle, et de comprendre au mieux la gestion temporelle de la micro-planification et de la formulation en décrivant de manière fine le contenu linguistique des unités (d'un point de vue sémantique, syntaxique, lexical), en distinguant les différentes variables chronométriques d'écriture (pauses initiales d'unités, pauses inter-mots, pauses intra-mots, temps d'exécution), et en effectuant les analyses sur des groupes de clauses relativement homogènes d'un point de vue syntaxique (clauses matrices, subordonnées, autonomes). Plusieurs perspectives pourraient

être envisagées pour poursuivre ce travail. À partir du même corpus, il serait intéressant d'établir des profils d'ordre « linguistique », c'est-à-dire d'identifier des profils de scripteurs différant par exemple quant au schéma textuel, au liage syntaxique, aux usages lexicaux, puis de mettre en lien les profils « chronométriques » et « linguistiques » avec la qualité des textes, afin de caractériser à différents points de vue ce qui permet d'obtenir un « bon texte ». Par ailleurs, afin de poursuivre les analyses menées grâce aux données chronométriques, il serait envisageable de s'intéresser au niveau « infra-clausal », en intégrant la gestion de la production des syntagmes au sein des clauses, puis des mots au sein des syntagmes, rejoignant ainsi le travail entrepris par Maggio (2013) au niveau du mot. Enfin, de plus en plus de productions écrites étant effectuées sur ordinateur, surtout pour les adultes, une comparaison de la gestion des unités de production lors d'une rédaction manuscrite et lors d'une rédaction sur ordinateur permettrait de faire ressortir ce qui est dû au scripteur, au medium d'écriture, et ce qui parait plus général.

Enfin, nous avons intégré dans nos analyse l'étude des révisions, non pas pour les étudier « en soi » mais pour mieux comprendre la gestion temporelle des processus rédactionnels. Or, plusieurs éléments, dont la présence fréquente de « reprises esthétiques » en frontière d'unité linguistique ou avant des sessions de révision, nous ont interrogés sur le rôle de la relecture pendant la production de texte. Il serait alors très intéressant de pouvoir enregistrer de manière précise les mouvements oculaires pendant la production. Cela permettrait notamment d'explorer dans quelle mesure la relecture peut aider à la planification du texte à écrire, et plus particulièrement comment elle permet de lier les clauses entre elles, ce liage étant à la fois indispensable et difficile dans le cadre de la production textuelle.

Références

Abelson, R. P. (1981). Psychological status of the script concept. American Psychologist, 36(7), 715–729.

Adam, J.-M. (1985). *Le texte narratif: Traité d'analyse textuelle des récits.* Nathan.

Adam, J.-M. (1991). Langue et littérature: analyses pragmatiques et textuelles. Hachette.

Adam, J.-M. (1993). Le texte et ses composantes. Théorie d'ensemble des plans d'organisation. SEMEN, 8, 11–40.

Adam, J.-M. (2011a). La linguistique textuelle (3e éd.). Paris: Armand Colin.

Adam, J.-M. (2011b). The narrative sequence : history of a concept and a research area. The sequence in classical French narratology, (1968), 39–43.

Akiguet, S., & Piolat, A. (1996). Insertion of connectives by 9- to 11-year-old children in an argumentative text. Argumentation, 10, 253–270.

Akinci, M.-A. (2005). La complexité syntaxique dans les textes écrits en français: étude chez des bilingues et monolingues.

Alamargot, D. (2005). Le rôle de la lecture au cours de l'écriture : ce que nous indiquent les mouvements oculaires du rédacteur. Rééducation Orthophonique, 223, 189–201.

Alamargot, D., Caporossi, G., Chesnet, D., & Ros, C. (2011). What makes a skilled writer? Working memory and audience awareness during text composition. Learning and Individual Differences, 21, 505–516.

Alamargot, D., & Chanquoy, L. (2001). Through the models of writing. Studies in Writing (Vol. 9). Dordecht - Boston - London: Kluwer Academic Publishers.

Alamargot, D., & Chanquoy, L. (2002). Les modèles de rédaction de textes. In M. Fayol (Ed.), Production du langage (p. 45–65). Paris.

Alamargot, D., Chesnet, D., Dansac, C., & Ros, C. (2006). Eye and pen: A new device for studying reading during writing. Behavior Research Methods, 38(2), 287–299.

Alamargot, D., Dansac, C., Chesnet, D., & Fayol, M. (2007). Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition : Research and applications*. Amsterdam.

Alamargot, D., Dansac, C., Ros, C., & Chuy, M. (2005). Rédiger un texte procédural à partir de sources: Relations entre l'empan de production écrite et l'activité oculaire du scripteur. In D. Alamargot, P. Terrier, & J. M. Cellier (Eds.), Production, compréhension et usage des écrits techniques au travail. (p. 51–68). Octares.

Alamargot, D., Flouret, L., Larocque, D., Caporossi, G., Pontart, V., Paduraru, C., … Fayol, M. (2015). Successful written subject–verb agreement: an online analysis of the procedure used by students in Grades 3, 5 and 12. Reading and Writing, 28(3), 291–312.

Alamargot, D., Lambert, E., & Chanquoy, L. (2005). La production écrite et ses relations avec la mémoire. *Approche Neuropsychologique Des Acquisitions de l'Enfant*, 17, 41–46.

Alamargot, D., & Lebrave, J.-L. (2010). The study of professional writing: A joint contribution from cognitive psychology and genetic criticism. European Psychologist, 15(1).

Alamargot, D., Plane, S., Lambert, E., & Chesnet, D. (2010). Using eye and pen movements to trace the development of writing expertise: case studies of a 7th, 9th and 12th grader, graduate student, and professional writer. Reading and Writing, 23(7), 853–888.

Allen, S., Skarabela, B., & Hughes, M. (2008). Using corpora to examine discourse effects in syntax. Trends in Language Acquisition Research, Corpora in Language Acquisition Research, 6, 99–137.

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: the role of grammatical units. Journal of Experimental Psychology. Learning, Memory, and Cognition, 33(4), 791–810.

Allum, P. H., & Wheeldon, L. R. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual–syntactic interface. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(5), 1240–1255.

Alves, R. A., Castro, S. L., & Olive, T. (2008). Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill. International Journal of Psychology, 43(6), 969–979.

Alves, R. A., de Sousa, L., & Strömqvist, S. (2007). Influence of typing skill on pause –execution cycles in written composition. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), Writing and *cognition : Research and applications* (p. 55–65). Amsterdam: Elsevier.

Aparaci, M. (2010). El desarrollo de la conectividad discursiva en diferentes féneros y modalidades de produccion. Universidad de Barcelona, Spain.

Ariel, M. (1990). Accessing Noun Phrase Antecedents. London: Routledge.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. P. Spooren (Eds.), Text Representation: Linguistic and Psycholinguistic Aspects.

Aurnague, M. (2008). Qu'est-ce qu'un verbe de déplacement?: critères spatiaux pour une classification des verbes de déplacement intransitifs du français. In Congres Mondial De Linguistique Française (p. 1905–1917). Paris.

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: reflections on procedures and measures. Written Communication, 29(3), 1–32.

Baayen, R. H., Davidson, D. J., & Bates, D. N. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59, 390–412.

Baccino, T., & Pynte, J. (1998). Spatial encoding and referential processing during reading. European Psychologist, Vol. 3(No.1), 51–61.

Baddeley, A. (2000). The episodic buffer : a new component of working memory ? Trends in Cognitive Sciences, 4(11), 417–423.

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory, vol. 8 (p. 47–89). New York: Academic Press.

Barjolle, É. (2001). Une gêne pratique à l'égard des manuels. *Le Français Aujourd'hui*, 135(4), 31–43.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. Language and Speech, 22(3), 201–211.

Beauvais, C., Olive, T., & Passerault, J.-M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing processes. Journal of Educational Psychology, 103(2), 415–428.

Béguelin, M.-J. (2002). Clause, période ou autre ? La phrase graphique et la question des niveaux d'analyse. Verbum, XXIV(1–2), 85–107.

Benoit, J., & Fayol, M. (1989). Le développement de la catégorisation des types de textes. Pratiques, 62, 71–85.

Benveniste, E. (1966). Problèmes de linguistique générale. Paris: Gallimard.

Bereiter, C., & Scardamalia, M. (1987). The psychology of written composition. London.

Berman, R. A. (2004a). Introduction : Developing discourse stance in different text types and languages. Journal of Pragmatics, 37, 105–124.

Berman, R. A. (2004b). Language development across childhood and adolescence. (R. A. Berman, Ed.). Amsterdam & Philadelphia: John Benjamins B.V.

Berman, R. A. (2005). Introduction: Developing discourse stance in different text types and languages. Journal of Pragmatics, 37(2), 105–124.

Berman, R. A. (2008). The psycholinguistics of text construction. Journal of Child Language, 35, 1–37.

Berman, R. A., & Katzenberger, I. (2004). Form and Function in Introducing Narrative and Expository Texts : A Developmental Perspective. Discourse Processes, 38(1), 57–94.

Berman, R. A., & Nir, B. (2009). Clause-packaging in narratives: A crosslinguistic developmental study. In Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan I. Slobin (p. 149–162). New York: Taylor & Francis.

Berman, R. A., & Nir-Sagiv, B. (2004). Linguistic indicators of inter-genre differentiation in later language development. Journal of Child Language, 31(2), 339–380.

Berman, R. A., & Ravid, D. (2009). Becoming a literate language user: Oral and written text construction across adolescence. In D. Olson & N. Torrance (Eds.), Cambridge handbook of literacy (p. 92–111). Cambridge University Press.

Berman, R. A., & Slobin, D. I. (1994). Relating Events in Narrative: a Crosslinguistic Developmental Study. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berman, R. A., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. Written Language & Literacy, 5(1), 1–43.

Berninger, V. W., Fuller, F., & Whitaker, D. (1996). A Process Model of Writing Development Across the Life Span. Educational Psychology Review, 8(3), 193–218.

Berninger, V. W., Nagy, W., & Beers, S. (2011). Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: contributions of syntax and transcription to translation. Reading and Writing, 24(2), 151–182.

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's Model of skilled writing to explain beginning and developing writing. In E. C. Butterfield (Ed.), *Children's writing : toward* a process theory of the development of skilled writing (p. 57–81). London and Greenwich: JAI Press Inc.

Berninger, V. W., Vaughan, K. B., Abbott, R. D., Rogan, L. W., Brooks, A., Reed, E., & Graham, S. (1997). Treatment of handwriting problems in beginning writers: transfer from handwriting to composition. Journal of Educational Psychology, 89(4), 652–666.

Berrendonner, A. (1993). La phrase et les articulations du discours. Le Français Dans Le Monde, 20–26.

Berrendonner, A. (2002). Les deux syntaxes. Verbum, XXIV(1–2), 23–36.

Bertrand, C., & Dumoulin, C. (2013). *La planification sémantique dans la production de texte : Analyse en temps réel des nouvelles informations dans le cadre de récits d'adolescents en classe de 5ème* et de 3ème. Mémoire d'orthophonie, Université Claude Bernard Lyon 1.

Bestgen, Y. (2012). Évaluation automatique de textes et cohésion lexicale. Discours, 11.

Bestgen, Y., & Costermans, J. (1997). Temporal markers of narrative structure: Studies in production. In J. Costermans & M. Fayol (Eds.), Processing Interclausal Relationships: Studies in the Production and Comprehension of Text (p. 201–217). Mahwah, NJ: Lawrence Erlbaum Associates.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), Handbook of Psycholinguistics (p. 945–978). San Diego: Academic Press, California.

Boehm, D. C. (1993). Mozartians, Beethovians, and the teaching of writing. Quarterly of the National Writing Project and the Center for the Study of Writing and Literacy, 15(2), 15–18.

Bonin, P., Fayol, M., & Chalard, M. (2001). Age of acquisition and word frequency in written picture naming. The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 54(2), 469–489.

Boomer, D. S. (1965). Hesitation and grammatical encoding. Language and Speech, 8(3), 148–158.

Boré, C., & Bosredon, C. (2013). La phrase selon les brouillons : un trajet entre l'oral et l'écrit. Le *Français Aujourd'hui*, (181), 13–24.

Boscolo, P. (1990). The construction of expository text. First Language, 10(1990), 217–230.

Bosredon, C. (2014). Représentations de l'écriture au cycle 3: une enquête croisée auprès d'enseignants et d'élèves. Pratiques, 161, 30–46.

Bourdin, B. (2002). Apprentissage de la gestion de la production et contraintes de capacité. In M. Fayol (Ed.), Production du langage (Lavoisier, p. 149–170). Paris.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. International Journal of Psychology, 29(5), 591–620.

Bourdin, B., & Fayol, M. (2002). Even in adults, written production is still more costly than oral production. International Journal of Psychology, 37(4), 219–227.

Bourdin, B., Fayol, M., & Darciaux, S. (1996). The comparison of oral and written modes on adults' and children's narrative recall. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), Theories, models and methodology in writing research (Amsterdam, p. 159–169). Amsterdam.

Breetvelt, I., van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality : when and how? Cognition and Instruction, 12(2), 103–123.

Brown, J. S., McDonald, J. L., Brown, T. L., & Carr, T. H. (1988). Adapting to processing demands in discourse production: the case of handwriting. Journal of Experimental Psychology: Human Perception and Performance, 14(1), 45–59.

Butterfield, E. C., Hacker, D. J., & Albertson, L. R. (1996). Environmental, cognitive, and metacognitive influences on text revision: Assessing the evidence. Educational Psychology Review, 8(3), 239–297.

Butterworth, B. L. (1975). Hesitation and semantic planning in speech. Journal of Psycholinguistic Research, 4(1), 75–87.

Caramazza, A. (1997). How many levels of processing are there in lexical access? Cognitive Neuropsychology, 14(1), 177–208.

Chafe, W. (1994). Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing. Discourse. Chicago: University of Chicago Press.

Chanfrault-Duchet, M.-F. (2001). La phrase au lycée : enjeux didactiques. *Le Français Aujourd'hui*, 135(4), 52–63.

Chanquoy, L. (2001). How to make it easier for children to revise their writing: a study of text revision from 3rd to 5th grades. The British Journal of Educational Psychology, 71(71), 15–41.

Chanquoy, L., & Alamargot, D. (2002). Mémoire de travail et rédaction de textes: Evolution des modèles et bilan des premiers travaux. *L'Année Psychologique*, 102, 363–398.

Chanquoy, L., & Alamargot, D. (2003). Mise en place et développement des traitements rédactionnels: le rôle de la mémoire de travail. *Le Langage et L'homme*, 38(2), 171–190.

Chanquoy, L., & Fayol, M. (1995). Analyse de l'évolution et de l' utilisation de la ponctuation et des connecteurs dans deux types de texte. Etude longitudinale du CP au CE2. Enfance, 48(2), 227–241.

Chanquoy, L., Foulin, J. N., & Fayol, M. (1990). Temporal management of short text writing by children and adults. Cahiers de Psychologie Cognitive, 10(5), 513–540.

Chanquoy, L., Foulin, J. N., & Fayol, M. (1996). Writing in adults: A real-time approach. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), Theories, models and methodology in writing research (Amsterdam, p. 36–43). Amsterdam.

Charolles, M. (1978). Introduction aux problèmes de la cohérence des textes. Langue Française, 38(1), 7–41.

Charolles, M. (2011). Cohérence et cohésion du discours. In K. Hölker & C. Marello (Eds.), Dimensionen der Analyse Texten und Diskursivent (p. 153–173).

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: generating text in L1 and L2. Written Communication, 18(1), 80–98.

Chenu, F., Pellegrino, F., Jisa, H., & Fayol, M. (2014). Interword and intraword pause threshold in writing. Frontiers in Psychology, 5, 1–7.

Chesnet, D., & Alamargot, D. (2005). Analyse en temps réel des activités oculaires et grapho-motrices du scripteur : intérêt du dispositif « Eye and Pen ». *L'année Psychologique*, 105(3), 477–520.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), Eye movement research: A window on mind and brain (p. 341–372). Oxford: Elsevier Ldt.

Coirier, P., Gaonac'h, D., & Passerault, J.-M. (1996). Psycholinguistique textuelle: une approche cognitive de la compréhension et de la production des textes. Paris.

Colognesi, S., & Deschepper, C. (2014). Le savoir des élèves en ponctuation. *Le Français Aujourd'hui*, 187(4), 43–55.

Combettes, B. (1986). Coréférence et connexité thématique dans le discours. In M. Charolles, J. S. Petöfi, & E. Sözer (Eds.), Research in Text Connexity and Text Coherence (p. 101–124). Helmut Buske Verlag Hamburg.

Connelly, V., Dockrell, J. E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. Educational Psychology, 25(1), 97–104.

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment. Written Communication, 29(3), 278–302.

Cook, M., Smith, J., & Lalljee, M. G. (1974). Filled pauses and syntactic complexity. Language and Speech, 17(1), 11–16.

Creissels, D. (2006). Syntaxe générale - *Une introduction typologique, 2 : La phrase* (Lavoisier). Paris: Hermès.

Crossley, S. A., & McNamara, D. S. (2010). Cohesion, Coherence, and Expert Evaluations of Writing Proficiency. In S. Ohlsson & R. Catrambone (Eds.), 32nd Annual Conference of the Cognitive Science Society (p. 984–989).

Crossley, S. A., & McNamara, D. S. (2011a). Text coherence and judgments of essay quality: Models of quality and coherence. In Proceedings of the 29th Annual Conference of the Cognitive Science Society (p. 1236–1241).

Crossley, S. A., & McNamara, D. S. (2011b). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. International Journal of Continuing Engineering Education and Life-Long Learning, 21(2/3), 170–191.

Daiute, C. A. (1984). Performance limits on writers. New Directions in Composition Research, 205–224.

David, J. (1994). La réécriture au confluent des approches linguiqtique, psychologique et didactique. Repères: Recherches En Didactique Du Français Langue Maternelle, 10, 3–13.

Degand, L., & Simon, A. C. (2009). On identifying basic discourse units in speech: theoretical and empirical issues. Discours. Revue de linguistique, psycholinguistique et informatique, 4, 1–19.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93(3), 283–321.

Donovan, C. A., & Smolkin, L. B. (2006). Children's understanding of genre and writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of Writing Research. New York and London: The Guilford Press.

Doquet, C. (1995). Le temps d'écrire : stratégies d'écriture et chronologie des évènements dans des processus d'écriture d'élèves de CM2. Repères, 11, 59–72.

Doquet, C. (2003). Écriture et traitement de texte à l'école élémentaire: modes d'analyse et pistes de travail. Langage et Société, 103, 11–29.

Doquet, C. (2006). L'objet insaisissable : quelques considérations sur l'écriture sur traitement de texte. Genesis, 27, 35–44.

Doquet, C., & Leblay, C. (2014). Temporalité de l'écriture et génétique textuelle : Vers un autre métalangage ? In 4è Congrès Mondial de Linguistique Française (p. 2767–2781).

Doquet-Lacoste, C. (2004). Indices et traces de l'activité métadiscursive des scripteurs : aspects de la réécriture. *Le Français Aujourd'hui*, 144(1), 33–41

Doquet-Lacoste, C. (2006). L'écriture débutante. Mise en texte et mise en graphie dans l'écriture sur traitement de texte à l'école. Langages, 164(4), 43–56.

Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-tracking and keylogging data. Journal of Writing Research, 5(1), 133–158.

Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.-B., & Mével, J.-P. (2007). Grand dictionnaire de linguistique et des sciences du langage.

Dufour, M.-P. (2014). *Des traités de ponctuation à la classe de français : dictatisation d'un objet de* savoir.

Epting, L. K., Gallena, E. M., Hicks, S. A., Palmer, E. N., & Weisberg, T. (2013). Read and think before you write: Prewriting time and level of print exposure as factors in writing and revision. Journal of Writing Research, 4(3), 239–259.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. Psychological Review, 102(2),

Escarabajal, M. C., & Kekenbosch, C. (1998). L'activité de catégorisation de substantifs, de verbes et de dérivés verbaux. Langages, 132, 69–86.

Fabre, C. (1991). Les brouillons d'écoliers ou l'entrée dans l'écriture. Actes de Lecture, 35, 1–11.

Fabre-Cols, C. (2004). Les brouillons et l'école : ce qu'a changé la critique génétique. Le Français *Aujourd'hui*, 144(1), 18–24.

Faigley, L., & Witte, S. P. (1981). Analyzing revision. College Composition and Communication, 32(4), 400–414.

Fauchart, N. (2001). Le travail de la phrase dans une grammaire pour écrire. *Le Français Aujourd'hui*, 135(4), 64–71.

Favart, M. (2005). Les marques de cohésion : leur rôle fonctionnel dans l'acquisition de la production écrite de texte. Psychologie Française, 50, 305–322.

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, 155(3), 51–68.

Favart, M., & Olive, T. (2005). Modèles et méthodes d'étude de la production écrite. Psychologie Francaise, 50(3), 273–285.

Favart, M., & Passerault, J.-M. (1996). Functionality of cohesion devices in the management of local and global coherence: Two studies in children's written production of narratives. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), Theories, models and methodology in writing research (p. 349–365). Amsterdam.

Favart, M., & Passerault, J.-M. (1999). Aspects textuels du fonctionnement et du développement des connecteurs : approche en production. *L'année Psychologique*, 99(1), 149–173.

Fayol, M. (1985). *Le récit et sa construction : une approche de psychologie cognitive.* Delachaux & Niestle.

Fayol, M. (1986). Les connecteurs dans les récits écrits. Pratiques, 49, 101–113.

Fayol, M. (1987). Vers une psycholinguistique textuelle génétique : l'acquisition du récit. In G. Pieraut - Le Bionec (Ed.), Connaître et le dire (Mardaga). Bruxelles.

Fayol, M. (1989). Une approche psycholinguistique de la ponctuation . Etude en production et en compréhension. Langue Française, 81, 21–39.

Fayol, M. (1990). La production de textes écrits. Introduction à l'approche cognitive. In R. Guibert & D. Jacobi (Eds.), *Les adultes et l'écriture* (p. 21–30).

Fayol, M. (1996). La production d'écrits narratifs: approche de psycholinguistique textuelle chez l'enfant et l'adulte. In J. David & S. Plane (Eds.), *L'apprentissage de l'écriture de l'école au* collège (p. 9–36). Paris: Presses Universitaires de France.

Fayol, M. (1997). Des idées au texte. Psychologie cognitive de la production verbale, orale et écrite. Paris: Presses Universitaires de France.

Fayol, M. (2005). Écrire, orthographier et rédiger des textes. Psychologie Française, 50, 391–402.

Fayol, M., Carré, M., & Simon-Thibult, L. (2014). Enseigner la ponctuation : comment et avec quels effets ? *Le Français Aujourd'hui*, 187(4), 31–40.

Fayol, M., & Heurley, L. (1995). Des modèles de production du langage à l'étude du fonctionnement du scripteur, enfant et adulte. *La Production de Textes: Vers Un Modèle D'enseignement de L'écriture*.

Fayol, M., & Lété, B. (2012). Contributions of online studies to understanding translation from ideas to written text. In M. Fayol, D. Alamargot, & V. W. Berninger (Eds.), Translation of Thought to Written Text While Composing (p. 289–314).

Ferrand, L. (2002). Les modèles de la production de la parole. In M. Fayol (Ed.), Production du langage (Lavoisier, p. 27–44). Paris.

Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. Instructional Science, 26(6), 473–488.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. Journal of Memory and Language, 30, 210–233.

Fiorin, J. L. (1998). Temps : entre la langue et le discours. Linx.

Fleischman, S. (1985). Discourse functions of tense-aspect oppositions in narrative: toward a theory of grounding. Linguistics, 23(6), 851–882.

Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In W. Gregg & L. Steinberg (Eds.), Cognitive processes in writing (p. 31–50). Hillsdale, NJ: Lawrence Erlbaum Associates.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. College Composition and Communication, 32(4), 365–387.

Foltz, P. W., & Landauer, T. K. (2003). Potential applications of latent semantic analysis to writing analysis. In T. Olive & C. M. Levy (Eds.), Contemporary Tools and Techniques for Studying Writing (Vol. 10). Dordrecht, Boston, London: Kluwer.

Fortier, G., & Préfontaine, C. (1994). Pauses, relecture et processus d'écriture. Revue Des Sciences de *L'éducation*, 20(2), 203–220.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. Applied Linguistics, 21(3), 354–375.

Foulin, J. N. (1995). Pauses et débits: les indicateurs temporels de la production écrite. *L'année* Psychologique, 95(3), 483–504.

Foulin, J. N. (1998). To what extent does pause location predict pause duration in adults' and children's writing? Cahiers de Psychologie Cognitive, 17(3), 601–620.

Foulin, J. N., Chanquoy, L., & Fayol, M. (1989). Approche en temps réel de la production des connecteurs et de la ponctuation : vers un modèle procédural de la composition écrite. Langue Française, 81, 5–20.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. Language, 47(1), 27–52.

Gaatone, D. (2008). Le prédicat: pour quoi faire? LIDIL, 37, 45–60.

Galbraith, D. (1992). Conditions for discovery through writing. Instructional Science, 21(1–3), 45–71.

Galbraith, D. (1999). Writing as a Knowledge-constituting process. In M. Torrance & D. Galbraith (Eds.), Knowing What to Write. Conceptual Processes in Text Production (p. 139–160). Amsterdam: Amsterdam University Press.

Galbraith, D. (2009). Writing as discovery. Teaching and Learning Writing, 1(1), 5–26.

Garcia-Debanc, C., & Fayol, M. (2013). About the psycholinguistic models of the writing process for a didactics of written production. Repères, Hors-série, 2–15.

Garnham, A., Oakhill, J., & Johnson-Laird, P. N. (1982). Referential continuity and the coherence of discourse. Cognition, 11, 29–46.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. Developmental Psychology, 40(2), 177–190.

Gayraud, F. (2001). Le développement de la différenciation oral/écrit vu à travers le lexique. Thèse de doctorat, Université Lyon 2.

Gayraud, F., Jisa, H., & Viguie, A. (2001). Utilisation des outils cohésifs comme indice de sensibilité au registre: une étude développementale. Acquisition et Interaction En Langue Étrangère, 14, 3–24.

Gayraud, F., Jisa, H., & Viguié, A. (1999). The development of syntactic packaging in French children's written and spoken texts. In R. A. Aisenman (Ed.), Developing Literacy Across Genres, Modalities, and languages (p. 169–181). Tel Aviv University Press.

Gayraud, F., & Martinie, B. (2004). Complexité syntaxique et difficulté de traitement : l'exemple des relatives. In A. Rabatel (Ed.), Interactions orales en contexte didactique. Mieux (se) comprendre *pour mieux (se) parler et mieux (s')apprendre*. PUL.

Ghiglione, R., Bromberg, M., Friemel, E., Kekenbosch, C., & Verstiggel, J.-C. (1990). Prédications d'état, de déclaration et d'action: essai de classification en vue d'une application en analyse de contenu. Langages, 25(100), 81–100.

Givón, T. (1983). Topic coninuity in Discourse. Amesterdam / Philadelphia: John Benjamins.

Goldman, J.-P., François, T., Roekhaut, S., Simon, A. C., & others. (2010). Étude statistique de la durée pausale dans différents styles de parole. *Journées d'Etude Sur La Parole (JEP)*.

Goldman-Eisler, F. (1961). The distribution of pause durations in speech. Language and Speech, 4(4), 232–237.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. Language and Speech, 15(2), 103–113.

Graesser, A. C., & McNamara, D. S. (2004). Coh-Metrix: Analysis of text on cohesion and language. Behavior Research Methods, Instruments, & Computers, 36(2), 193–202.

Greene, J. O., & Cappella, J. N. (1986). Cognition and talk: The relationship of semantic units to temporal patterns of fluency in spontaneous speech. Language and Speech, 29(2), 141–157.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. Cognition, 82(1).

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. Psychonomic Bulletin & Review, 10(3), 603–609.

Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. Langages, 15(63), 7–52.

Gross, M. (1993). Les phrases figées en français. L Information Grammaticale, 59(1), 36–41.

Haiman, J., & Thompson, S. A. (1988). Clause combining in grammar and discourse. Amsterdam & Philadelphia: John Benjamins (Typological Studies in Language, 18).

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in english.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring New Information as a Process in Comprehension. Journal of Verbal Learning and Verbal Behavior, 13, 512–521.

Hawkins, P. R. (1971). The syntactic location of hesitation pauses. Language and Speech, 14(3), 277–288.

Hayes, J. R. (1991). *Peeking Out from Under the Blinders: Some Factors We Shouldn't Forget in* Studying Writing.

Hayes, J. R. (1995). Un nouveau modèle du processus d'écriture. In La production de textes: vers un *modèle d'enseignement de l'écriture*.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. The Science of Writing: Theories, Methods, Individual Differences, and Applications, (March), 1–27.

Hayes, J. R. (1998). Un nouveau cadre pour intégrer cognition et affect dans la rédaction. In A. Piolat (Ed.), La rédaction de textes (p. 51–101). Lausanne: Delachaux & Niestle.

Hayes, J. R. (2012). Modeling and remodeling writing. Written Communication, 29(3), 369–388.

Hayes, J. R., & Berninger, V. W. (2010). Relationships between idea generation and transcription: How act of writing shapes what children write. In C. Braverman, R. Krut, K. Lunsford, S. McLeod, S. Null, P. Rogers, & A. Stansell (Eds.), Traditions of writing research (p. 166–180). New York: Taylor & Francis.

Hayes, J. R., & Flower, L. (1980). Identifying the Organization of Writing Processes. In L. Gregg & E. R. Steinberg (Eds.), Cognitive processes in writing (p. 3–30). Hillsdale, NJ: Erlbaum.

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), Reading, Writing, and language processes (p. 176–240). Cambridge, England: Cambridge University Press.

Hayes, J. R., & Nash, J. G. (1996). On the nature of planning in writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing : Theories, Methods, Individual Differences and Applicatio*ns (p. 29–55).

Heurley, L. (2006). La révision de texte: L'approche de la psychologie cognitive. Langages, 164(4), 10–25.

Holmes, V. M. (1984). Sentence planning in a story continuation task. Language and Speech, 27(2), 115–134.

Hopper, P. J., & Thompson, S. A. (2008). Projectability and Clause Combining in Interaction. In R. Laury (Ed.), Crosslinguistic Studies of Clause Combining: The Multifunctionality of Conjunctions (p. 99–124). Amsterdam: John Benjamins.

Hunt, K. W. (1965). Grammatical structures written at three grade levels. In National Council of Teachers of English, III.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. Monographs of the Society for Research in Child Development, 35(1), 67.

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye Movement Measures to Study Global Text Processing. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 313–334.

Immonen, S. (2006). Translation as a writing process: Pauses in translation versus monolingual text production. Target: International Journal of Translation Studies, 18(2), 313–336.

Immonen, S., & Mäkisalo, J. (2010). Pauses reflecting the processing of syntactic units in monolingual text production and translation. Hermes–Journal of Language and Communication Studies, 44, 45–61.

Jaffré, J.-P. (2014). Postface. *Le Français Aujourd'hui*, 187(4), 129.

Jakobsen, A. L. (2006). Research methods in translation—Translog. In K. P. H. Sullivan & E. Lindgren (Eds.), Computer keystroke logging and writing: Methods and applications (p. 95–105). Amesterdam: Elsevier.

Janssen, D., van Waes, L., & van den Bergh, H. (1996). Effects of thinking aloud on writing process. The Science of Writing: Theories, Methods, Individual Differences, and Applications, (December 2007), 233–250.

Jeanneret, Y. (2010). Complexité de la notion de trace. De la traque au tracé. In B. Galinon-Mélénec (Ed.), *L'homme trace. Perspectives anthropologiques des traces contemporaines* (p. 59–86). Paris: CNRS.

Jisa, H. (2000). Increasing cohesion in narratives: a developmental study of maintaining and reintroducing subjects in French. Linguistics, 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. *Cahiers d'Acquisition et de Pathologie* Du Langage, 24, 33–56.

Jisa, H., Reilly, J., Verhoeven, L., Baruch, E., & Rosado, E. (2002). Passive vocie constructions in written texts. Written Language & Literacy, 5(2), 163–181.

Johansson, R., Johansson, V., Wengelin, Å., & Holmqvist, K. (2012). Reading during text production. In Current Trends in European Writing Research (p. 294–298).

Johansson, V. (2009). Developmental aspects of text production in writing and speech. Thèse de doctorat, Université de Lund.

Johansson, V., Gustafsson, P., & Wengelin, Å. (2015). Revising and pausing in relation to syntactic development. In EARLI 2015: Towards a reflective society. Synergies between learning, teaching and research.

Johansson, V., Johansson, R., Wengelin, Å., & Holmqvist, K. (2009). Individual strategies of pauses and reading during writing. In 13th Biennial Conference for Research on Learning and Instruction.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. Psychological Review, 87(4), 329–354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99(1), 122–149.

Katzenberger, I. (2004). The development of clause packaging in spoken and written texts. Journal of Pragmatics, 36(10), 1921–1948.

Katzenberger, I. (2005). The super-structure of written expository texts – a developmental perspective. In D. Ravid & H. B. Z. Shyldkrot (Eds.), Perspectives on language and language development. Essays in Honor of Ruth A. Berman (p. 327–336). Dordrecht, Boston, London: Kluwer Academic Publishers.

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. Research in the Teaching of English, 121–140.

Kekenbosch, C., Meunier, J.-M., Richard, J.-F., Desclés, J.-P., & Flageul, V. (1998). Sémantique cognitive de l'action : 2 . étude expérimentale de la catégorisation des verbes d'action. Langages, 132, 48–68.

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. Memory & Cognition, 15(3), 256–266.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), The *Science of Writing : Theories, Methods, Individual Differences and Applications* (p. 57–71). New Jersey.

Kellogg, R. T. (1999). Components of working memory in text production. In M. Torrance & G. C. Jeffrey (Eds.), The cognitive demands of writing: processing capacity and working memory effects in text production (p. 43–61). Amsterdam.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. Journal of Writing Research, 1(1), 1–26.

Kellogg, R. T., Turner, C. E., Whiteford, A. P., & Mertens, A. (2016). The role of working memory in planning and generating written sentences. Journal of Writing Research, 7(3), 397–416.

Kintsch, W., & Van Dijk, T. A. (1975). Comment on se rappelle et on résume des histoires. Langages, 9(40), 98–116.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85(5), 363–394.

Kleiber, G. (2003). Faut-il dire adieu à la phrase ? *L'information Grammaticale*, (98), 17–22.

Koch, P. (1995). Subordination, intégration syntaxique et oralité. In H. L. Andersen & G. Skytte (Eds.), La subordination dans les langues romanes.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. Journal of Memory and Language, 66(1), 143–162.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. Cognitive Psychology, 73, 1–40.

Krämer, S. (2010). Qu'est-ce donc qu'une trace, et quelle est sa fonction épistémologique ? État des lieux. Trivium, 10, 11–27.

Labov, W., & Waletzky, J. (1967). Narrative analysis: oral versions of personal experience. In J. Helm (Ed.), Essays on the verbal and visual arts (p. 12–44). Seattle & London: University of Washington Press.

Le Goffic, P. (2001). La phrase « revisitée ». *Le Français Aujourd'hui*, 135(4), 96.

Le Goffic, P. (2006). Phrase, séquence, période. Recherches ACLIF: Actes Du Séminaire de Didactique Universitaire, (3), 90–115.

Leblay, C. (2009). La question du "déjà écrit" dans le processus d'écriture observé en temps réel. Modèles Linguistiques, 59, 153–176.

Lebrave, J. (2009). Manuscrits de travail et linguistique de la production écrite. Modèles Linguistiques, 59, 13–21.

Lebrave, J., & Grésillon, A. (2009). Linguistique et génétique des textes : un décalogue. [En ligne], Mis en ligne le 23 mars 2009. Disponible Sur: http://lodel. item. ens. fr/item/index. php., 1–13.

Lebrave, J.-L. (2002). La critique génétique et l'étude des processus d'écriture littéraires. In M. Fayol (Ed.), Production du langage (Lavoisier, p. 251–262). Paris.

Legallois, D. (2006). Présentation générale. Le texte et le problème de son et ses unités : propositions pour une déclinaison. Langage et Société, 3, 3–9.

Leijten, M., & van Waes, L. (2006). Inputlog: New Perspectives on the Logging of On-Line Writing. In K. P. H. Sullivan & E. Lindgren (Eds.), Computer Keystroke Logging: Methods and Applications (p. 73–94). Oxford: Elsevier.

Leijten, M., van Waes, L., & van Horenbeeck, E. (2015). Analyzing writing process data - A linguistic perspective. In G. Cislaru (Ed.), Writing(s) at the Crossroads - The process-product interface (p. 277–302). Amsterdam & Philadelphia: John Benjamins.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). Manulex: A grade-level lexical database from French elementary-school readers. Behavior Research Methods, Instruments, & Computers, 36, 156–166.

Leuwers, C. (2002). La production de phrases. In M. Fayol (Ed.), Production du langage (Lavoisier, p. 107–130). Paris.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge & London: MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), The neurocognition of language (p. 94–122). Oxford: Oxford University Press.

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. Crossing the Boundaries in Linguistics, 221–252.

Lindgren, E., & Sullivan, K. P. H. (2006a). Analysing online revision. In K. P. H. Sullivan & E. Lindgren (Eds.), Computer Keystroke Logging: Methods and Applications (Vol. 18, p. 157–188). Amsterdam.

Lindgren, E., & Sullivan, K. P. H. (2006b). Writing and the analysis of revision: An overview. In K. P. H. Sullivan & E. Lindgren (Eds.), Computer Keystroke Logging: Methods and Applications (Vol. 18, p. 31–44). Amsterdam: Elsevier Ldt.

Maggio, S. (2013). *La dynamique de la production verbale écrite : le traitement du mot dan*s le texte et les syntagmes. Thèse de doctorat, Université Blaise Pascal - Clermont Ferrand.

Maggio, S., Lété, B., Chenu, F., Jisa, H., & Fayol, M. (2012). Tracking the mind during writing: immediacy, delayed, and anticipatory effects on pauses and writing rate. Reading and Writing, 25(9), 2131–2151.

Maillard, M. (2008). Le prédicat : Comment sortir de la Tour de Babel? LIDIL, 37, 23–43.

Maingueneau, D. (1993). *Le contexte de l'oeuvre littéraire : énonciation , écrivain , société.* Paris: Dunod.

Marti, M. (1999). La clôture narrative, propositions théoriques d'analyse. Narratologie, 134–154.

Martin, F. (2007). Predicats statifs , causatifs et resultatifs en discours Semantique des adjectifs evaluatifs et des verbes psychologiques. Université Libre de Bruxelles.

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. Cognition, 116(2), 177–192.

Matsuhashi, A. (1981). Pausing and Planning: The Tempo of Written Discourse Production. Research in the Teaching of English, 15(2), 113–134.

Matsuhashi, A. (1982). Explorations in the real-time production of written discourse. In M. Nystrand (Ed.), What writers know: The language, process, and structure of written discourse (p. 269–290). New York: Academic Press.

Matthiessen, C., & Thompson, S. A. (1988). The structure of discourse and "subordination." In Clause combining in grammar and discourse (p. 275–329).

Mazur-Palandre, A. (2007). Later language development: syntactic packaging in written and spoken French. In The second Oxford Postgraduate Conference (p. 173–180).

Mazur-Palandre, A. (2009). *Le flux de l'information, aspects syntaxiques et discursifs* - Une étude fonctionnaliste et développementale. Thèse de doctorat, Université Lyon 2.

Mazur-Palandre, A. (2015). Overcoming preferred argument structure in written French: Development, modality, text type. Written Language & Literacy, 18, 25–55.

McCulley, G. A. (1985). Writing Quality, Coherence, and Cohesion. Research in the Teaching of English, 19(3), 269–282.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. Educational Psychology Review, 8(3), 299–325.

McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of Writing Research (p. 115–130). New York and London: The Guilford Press.

McCutchen, D. (2011). From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill. Journal of Writing Research, 3(1), 51–68.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic Features of Writing Quality. Written Communication, 17(1), 57–86.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: results from picture–word interference experiments. Journal of Memory and Language, 35, 477–496.

Molendijk, A. (1985). Point référentiel et imparfait. Langue Française, 67, 78–94.

Molendijk, A. (1990). *Le Passé Simple et l'Imparfait : une approche Reichenbachienne*. Amsterdam / Atlanta: Rodopi.

Mouchon, S., Fayol, M., & Gombert, J. É. (1989). L'utilisation de quelques connecteurs dans des rappels de récits chez des enfants de 5 à 8 ans. *L'année Psychologique*, 89(4), 513–529.

Myhill, D. (2009). From talking to writing: Linguistic development in writing. British Journal of Educational Psychology, 2(6), 27–44.

Nippold, M. A. (2004). Research on later language development - International perspectives. In R. A. Berman (Ed.), Language Development across Childhood and Adolescence (p. 1–8). Amsterdam & Philadelphia: John Benjamins B.V.

Nir, B. (2008). Clause packages as constructions in developing narrative discourse. Tel Aviv University.

Nir, B., & Berman, R. A. (2010). Complex syntax as a window on contrastive rhetoric. Journal of Pragmatics, 42, 744–765.

O'Hare, F. (1973). Sentence Combining: Improving Student Writing Without Formal Grammar Instruction. NCTE Research Report, 15.

Olive, T. (2002). La gestion en temps réel de la production verbale: méthodes et données. In M. Fayol (Ed.), Production du langage. Paris: Hermès Science Publications.

Olive, T. (2004). Working Memory in Writing: Empirical Evidence From the Dual-Task Technique. European Psychologist, 9, 32–42.

Olive, T. (2011). L'analyse du décours temporel de la production écrite. In T. Ponchon & I. Laborde-Milaa (Eds.), Sciences du langage et nouvelles technologies (p. 61–68). Limoges: Lambert-Lucas.

Olive, T. (2014). Toward a Parallel and Cascading Model of the Wri ting System: A Review of Research on Writing Processes Coordination. Journal of Writing Research, 6(2), 173–194.

Olive, T., Alves, R. A., & Castro, S. L. (2009). Cognitive processes in writing during pause and execution periods. European Journal of Cognitive Psychology, 21(5), 758–785.

Olive, T., & Cislaru, G. (2015). Linguistic forms at the process-product interface : Analysing the linguistic content of bursts of production. In G. Cislaru (Ed.), Writing(s) at the Crossroads - The process-product interface (p. 99–123). Amsterdam & Philadelphia: John Benjamins.

Olive, T., Favart, M., Beauvais, C., & Beauvais, L. (2009). Children's cognitive effort and fluency in writing: Effects of genre and of handwriting automatisation. Learning and Instruction, 19, 229–308.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. Memory & Cognition, 30(4), 594–600.

Olive, T., Kellogg, R. T., & Piolat, A. (2002). The triple task technique for studying the process of writing. In T. Olive & C. M. Levy (Eds.), Contemporary tools and techniques for studying writing (p. 31–59).

Olive, T., Kellogg, R. T., & Piolat, A. (2008). Verbal, visual, and spatial working memory demands during text composition. Applied Psycholinguistics, 29, 669–687.

Oliveira, M. (2002a). Pausing strategies as means of information processing in spontaneous narratives. In Proceedings of the 1st International Conference on Speech Prosody.

Oliveira, M. (2002b). The role of pause occurrence and pause duration in the signaling of narrative structure. Advances in Natural Language Processing, 43–51.

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, 177(1), 113–128.

Paolacci, V., & Rossi Gensane, N. (2014). Ponctuation et écrits d'élèves : Une conception différente de la phrase pour enseigner la ponctuation autrement. Le Français Aujourd*'hui*, 187(4), 115–125.

Paolacci, V., & Rossi-Gensane, N. (2012). Quelles images de la phrase dans les écrits d'élèves de fin d'école primaire française ? Description linguistique et réponses didactiques aux difficultés des élèves. SHS Web of Conferences, 1, 341–359.

Passerault, J.-M. (1991). La ponctuation - Recherches en psychologie du langage. Pratiques, 70, 85–104.

Peverly, S. T., & Sumowski, J. F. (2012). What variables predict quality of text notes and are text notes related to performance on different types of tests? Applied Cognitive Psychology, 26(1), 104–117.

Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic parsing. In M. J. Traxler & M. A. Gernsbacher (Eds.), Handbook of Psycholinguistics (p. 455–503). Amsterdam: Academic Press.

Piolat, A. (1983). Localisation syntaxique des pauses et planification du discours. *L'année* Psychologique, 83(2), 377–394.

Piolat, A. (2004). Approche cognitive de l'activité rédactionnelle et de son acquisition. Le rôle de la mémoire de travail. Linx, 51, 55–74.

Piolat, A., & Olive, T. (2000). Comment étudier le coût et le déroulement de la rédaction de textes. La méthode de la triple tâche - Un bilan méthodologique. *L'Année Psychologique*, 100, 465–502.

Plane, S. (2003). Stratégies de réécriture et gestion des contraintes d'écriture par des élèves de l'école élémentaire : ce que nous apprennent des écrits d'enfants sur l'écriture. Rivisita Italiana de Psicolinguistica Applicatta Anno III/1, 57–77.

Plane, S. (2004). Quels modèles pour analyser la production d'écrit sur traitement de texte? Les contraintes comme outil d'analyse et d'intervention. Linx, 51, 1–12.

Plane, S., Alamargot, D., & Lebrave, J.-L. (2010). Temporalité de l'écriture et rôle du texte produit dans l'activité rédactionnelle. Langages, 177(1), 7–28.

Plane, S., Olive, T., & Alamargot, D. (2010). Présentation: Pour une approche pluridisciplinaire des contraintes de la production écrite. Langages, 177(1), 3–5.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), Radical pragmatics (p. 223–255). New York: Academic Press.

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. P. (2008). Assessing the microstructure of written language using a retelling paradigm. American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association, 17(2), 107–120.

Pynte, J., New, B., & Kennedy, A. (2008a). A multiple regression analysis of syntactic and semantic influences in reading normal text. Journal of Eye Movement Research, 2(1), 1–11.

Pynte, J., New, B., & Kennedy, A. (2008b). On-line contextual influences during reading normal text: a multiple-regression analysis. Vision Research, 48, 2172–2183.

Quinlan, T., Loncke, M., Leijten, M., & van Waes, L. (2012). Coordinating the Cognitive Processes of Writing : The Role of the Monitor. Written Communication, 29(3), 345–368.

Rabatel, A. (2004). *Argumenter en racontant : (re)lire et (ré)écrire les textes littéraires*. De Boeck Université.

Ramasawmy, N. (2004). *Conjunctive cohesion and relational coherence in student's compositions*. University of South Africa.

Ransdell, S., Levy, C. M., & Kellogg, R. T. (2002). Effects of secondary task demands on writing. L-1: Educational Studies in Language and Literature, 2, 141–163.

Ravid, D., & Cahana-Amitay, D. (2005). Verbal and nominal expressions in narrating conflict situations in Hebrew. Journal of Pragmatics, 37, 157–186.

Rey-Debove, J. (1981). Pour une lecture de la rature. Semiotica, 36(1–2), 1–32.

Riegel, M., Pellat, J.-C., & Rioul, R. (2002). Grammaire méthodique du français. PUF.

Rochester, S. R., & Gill, J. (1973). Production of complex sentences in monologues and dialogues. Journal of Verbal Learning and Verbal Behavior, 12(2), 203–210.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. Cognition, 64, 249–284.

Roeser, J., Torrance, M., & Baguley, T. (2015). Syntactic planning and lexical access in sentence production. In Proceedings of the Architectures and Mechanisms for Language Processing (AMLaP) Conference.

Rondelli, F. (2013). La phrase, segment textuel « de base » ? *Le Français Aujourd'hui*, 181(2), 71–81.

Salles, M. (2006). Cohésion-cohérence : accords et désaccords. Corela, 5.

Sanders, T., & Pander Maat, H. (2006). Cohesion and coherence : linguistic approaches. In K. Brown (Ed.), Encyclopedia of Language and Linguistics. Volume 2 (Elsevier, p. 591–595). Amsterdam.

Sanders, T., & Schilperoord, J. (2006). Text structure as a window on the cognition of writing : how text analysis provides insights in writing products ans writing processes. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of Writing Research (p. 386–402). New York and London: The Guilford Press.

Scardamalia, M., & Bereiter, C. (1983). The development of evaluative, diagnostic, and remedial capabilities in children's composing. The Psychology of Written Language: Developmental and Educational Perspective, 67–95.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In Toward a general theory of expertise (p. 172–194). Cambridge: Cambridge University Press.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals, and understanding. New Jersey: Lawrence Erlbaum Associates.

Schilperoord, J. (1996). The distribution of pause time in written text production. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), Theories, models and methodology in writing research (p. 21–35). Amsterdam: Amsterdam University Press.

Schilperoord, J. (2002). On the cognitive status of pauses in discourse production. In T. Olive & M. Levy (Eds.), Contemporary tools and techniques for studying writing (p. 61–87). Dordrecht, The Netherlands: Academic Publishers.

Schilperoord, J., & Sanders, T. (1999). How hierarchical text structure affects retrieval processes: Implications of pause and text analysis. In Knowing What to Write. Conceptual Processes in Text Production (p. 13–33).

Schnedecker, C., & Landragin, F. (2014). Les chaînes de référence: présentation. Langages, 3, 3–22.

Schneuwly, B. (1988). *Le langage écrit chez l'enfant*. Neuchâtel, Paris: Delachaux & Niestle.

Schneuwly, B., Rosat, M.-C., & Dolz, J. (1989). Les organisateurs textuels dans quatre types de textes écrits. Etude chez des élèves de 10, 12 et 14 ans. Langue Française, 81, 40–58.

Schriefers, H., Teruel, E., & Meinshausen, R. M. (1998). Producing simple sentences: Results from picture-word interference experiments. Journal of Memory and Language, 39(4), 609–632.

Serres, A. (2002). Quelle(s) problématique(s) de la trace? In Séminaire du CERCOR (p. 1–15).

Severinson Eklundh, K., & Kollberg, P. (2003). Emerging discourse structure: Computer-assisted episode analysis as a window to global revision in university students' writing. Journal of Pragmatics, 35(6), 869–891.

Smith, M., & Wheeldon, L. R. (1999). High level processing scope in spoken sentence production. Cognition, 73(3), 205–246.

Spelman Miller, K. (2002a). Units of production in writing: evidence of topic "framing"in on-line writing research. Reading.

Spelman Miller, K. (2002b). Units of Production in Writing: Towards a Discourse Perspective. Revista Canaria de Estudios Ingleses, 115–139.

Spooren, W., & Sanders, T. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. Journal of Pragmatics, 40(12), 2003–2026.

Steinberg, L. (2005). Cognitive and affective development in adolescence. Trends in Cognitive Sciences, 9(2), 69–74.

Strömqvist, S., Holmqvist, K., Johansson, V., Karlsson, H., & Wengelin, Å. (2006). What keystroke logging can reveal about writing. In K. P. H. Sullivan & E. Lindgren (Eds.), Computer Keystroke Logging: Methods and Applications (Vol. 18, p. 45–71). Amsterdam: Elsevier Ldt.

Strömqvist, S., & Malmsten, L. (1998). Scriptlog pro 1.04 – *user's manual*.

Sumner, E., Connelly, V., & Barnett, A. (2013). Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution. Reading and Writing, 26(6), 991–1008.

Tierney, R. J., & Mosenthal, J. H. (1983). Cohesion and Textual Coherence. Research in the Teaching of English, 17(3), 215–29.

Tolchinsky, L. (2004). The nature and scope of later language development. In R. A. Berman (Ed.), Language Development across Childhood and Adolescence (p. 233–248). Amsterdam & Philadelphia: John Benjamins B.V.

Torrance, M. (1996). Is writing expertise like other kinds of expertise ? In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), Theories, models and methodology in writing research (p. 3–9). Amsterdam: Amsterdam University Press.

Torrance, M. (2015). Understanding planning in text production. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of Writing Research (2nd edition). New York: Guildford Press.

Torrance, M., Johansson, R., Johansson, V., & Wengelin, Å. (2015). Reading during the composition of multi-sentence texts: an eye-movement study. Psychological Research.

Torrance, M., & Nottbusch, G. (2012). Written production of single words and simple sentences. In V. W. Berninger (Ed.), Past, present, and future contributions of cognitive writing research to cognitive psychology (p. 403–423). New York: Psychology Press.

Touratier, C. (1989). Récit et temps verbaux. L Information Grammaticale, 41(1), 3–5.

van den Eynde, K., & Mertens, P. (2010). Le dictionnaire de valence DICOVALENCE: manuel *d'utilisation.* (version 2.0 du 2010-06-25).

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. Human Movement Science, 10, 165–191.

van Hell, J. G., Verhoeven, L., Tak, M., & van Oosterhout, M. (2005). To take a stance: a developmental study of the use of pronouns and passives in spoken and written narrative and expository texts in Dutch. Journal of Pragmatics, 37(2), 239–273.

van Hell, J. G., Verhoeven, L., & van Bijsterverldt, L. M. (2008). Pause time patterns in writing narrative and expository texts by children and adults. Discourse Processes, 45, 406–427.

van Waes, L., Leijten, M., & Quinlan, T. (2010). Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity. Reading and Writing, 23, 803–834.

van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. Journal of Pragmatics, 35, 829–853.

Wallace, D. L., & Hayes, J. R. (1991). Redefining revision for freshmen. Research in the Teaching of English, 25(1), 54–66.

Weinrich, H. (1973). *Le Temps : le récit et le commentaire*. Paris: Ed. du Seuil.

Wengelin, Å. (2006). Examining pauses in writing: Theory, methods and empirical data. Computer Key-Stroke Logging and Writing: Methods and Applications, 18, 107–130.

Zhang, Q., & Damian, M. F. (2010). Impact of phonology on the generation of handwritten responses: Evidence from picture-word interference tasks. Memory & Cognition, 38(4), 519–528.

Annexes

Annexe 1 : Codage des clauses

1. Cas protoypique

Dans le cas prototypique, une clause est constituée d'un verbe conjugué (146).

(146) *il connaissait ces couloirs par cœur*, / pourtant il doutait [S23]

Cependant, certaines unités composées d'un verbe conjugué peuvent ne pas être des clauses, et à l'inverse certains segments peuvent être comptés comme clause, même s'ils ne contiennent pas de verbe conjugué, comme nous allons le détailler à présent.

2. Cas des verbes non fléchis

Les verbes dits « non-fléchis » ou « non-finis » apparaissent sous forme d'infinitif, de participe passé, de participe présent ou de gérondif. Ils sont rarement précédés de leur sujet, et il faut donc se référer à une autre clause (ou partie de clause) comprenant un verbe conjugué pour connaitre le sujet : le sujet du verbe non fléchi peut alors être coréférentiel à celui de ce verbe conjugué, ou non coréférentiel.

2.1. Sujet coréférentiel

2.1.1. Propositions infinitives servant à compléter un verbe : une seule clause

Dans les cas où un infinitif a été relevé après un verbe conjugué, avec une coréférence des sujets, il n'a été compté qu'une seule clause. Les verbes conjugués sont alors considérés comme des « verbes supports ». Parmi ces derniers, il y a tout d'abord les semi-auxiliaires « aller » et « venir de », servant à exprimer respectivement le futur proche et le passé proche, lorsqu'ils sont suivis d'un infinitif. Ainsi, (147) peut être paraphrasé par un conditionnel (148), qui sert de futur dans le passé.

(147) il savait / *qu'il* allait réussir [S23][81]

(148) il savait / *qu'il réussirait.*

Nous retrouvons ensuite les « modaux », répertoriés par Blanche-Benveniste (1990), et dont nous pouvons distinguer trois catégories : « épistémiques » (*va, doit, peut, risque de, a l'air de, a beau*), « de semblance » (avoir failli, sembler, paraitre) et « de processus » (commencer à, être en train de, cesser de, continuer à, finir de). Ces verbes modaux sont également appelés « non recteurs », étant compatibles

[81] Dans ces exemples, les prédicats concernés par la description (prédicats complexes, formes non finies) sont soulignés.

avec n'importe quel autre verbe, même impersonnel. Le dictionnaire électronique Dicovalence parle de « verbes adjoints » : ils se construisent avec un autre verbe infinitif, sans que cette proposition infinitive ne soit un terme de valence du premier ; les termes valenciels sont sélectionnés par le verbe à l'infinitif. Comme le montrent les paires suivantes, il est possible de supprimer le verbe non recteur, ce dernier ne servant qu'à modaliser le propos.

(149) et se dessina une grosse moustache noire / qui devrait faire illusion dans *l'obscurité de cette nuit sans lune* [S22]

(150) et se dessina une grosse moustache / *qui ferait illusion dans l'obscurité de cette* nuit sans lune

(151) il les salua avec un grognement indistinct / *qu'ils* ne semblèrent pas entendre [S22]

(152) il les salua avec un grognement indistinct / *qu'ils n'entendirent pas*

(153) il fait nuit noire dans ces conduits / et Dimitri commence à perdre la notion du temps [S08]

(154) il fait nuit noire dans ces conduits / et Dimitri perd la notion du temps

En outre, à ces verbes non recteurs s'ajoutent d'autres verbes qui, bien qu'ils ne servent pas seulement de modaux, sont associés au verbe à l'infinitif qui suit, du moment que les sujets sont coréférentiels, comme nous le montrent les exemples (155) et (156).

(155) Dimitri ne souhaitait pas tuer le gardien [S02]

(156) Sous le choc, le gardien ne pouvait pas se relever [S06]

Ainsi, dans le cas des « propositions infinitives complétives » identifiées dans les grammaires traditionnelles, nous n'avons compté qu'une clause lorsque le sujet des deux verbes était coréférentiel.

2.1.2. Propositions circonstancielles non finies : clauses séparées

Dans plusieurs cas, des verbes non fléchis ont été considérés comme appartenant à des clauses séparées, même s'ils partageaient le même sujet : c'est notamment le cas des nombreuses propositions circonstancielles. Ces verbes non-fléchis peuvent être des infinitifs (157), des participes passés (158), des participes présents (159) ou des gérondifs (160).

(157) Dimitri sortit / pour aller diner. [S10]

(158) il sortit de sa cellule / et marcha dans le couloir, / protégé par la pénombre [S02]

(159) Il se regarda dans le miroir, / pensant à sa liberté future [S13]

(160) Il referma la cellule / en laissant *le gardien à l'intérieur* [S15]

2.2. Sujets non coréférentiels

Lorsque le sujet d'un verbe non fléchi diffère de celui du verbe fléchi, deux clauses ont été comptées. C'est le cas des propositions complétives infinitives dont le verbe est contrôlé par l'objet ou l'oblique (161) ou qui sont précédées d'une tournure impersonnelle (162), ainsi que des subordonnées circonstancielles ou relatives ayant leur propre sujet (163) ou dépendant de l'objet ou de l'oblique (164).

(161) il demanda alors à son ami le gardien / de faire diversion [S05]

(162) il était beaucoup plus prudent / de continuer à pieds [S04]

(163) ses sensations revenues, / *il s'engagea dans le couloir* [S18]

(164) où Dimitri possédait une ferme / léguée par son parrain [S19]

Il y a cependant un cas où il n'a été compté qu'une seule clause, malgré le fait que les sujets soient différents : lors de l'utilisation de structures causatives, avec le verbe « faire » (165) ou « laisser » (166). Dans le Dicovalence, ces verbes sont appelés « verbes adjoints augmentants », le sujet du verbe prédicateur, ici à l'infinitif, apparaissant sous la forme d'un objet direct auprès du verbe adjoint.

(165) ensuite il fit rentrer le gardien [S06]

(166) et les laissa tomber par terre [S03]

3. Cas des verbes fléchis regroupés en une seule clause

Lorsque le scripteur a utilisé la tournure « il y a » pour marquer la durée (167) ou l'expression « c'est … qui/que » servant à cliver un groupe nominal (168), il n'a été compté qu'une seule clause. Le Dicovalence nomme ces verbes « verbes dispositifs », ne les considérant pas comme des prédicateurs.

(167) Dimitri repensa à son arrivée il y a de ça presq*u'un an* [S28]

(168) *C'est à ce moment*-là que Dimitri élança sa jambe vers la tête du gardien [S01]

Annexe 2 : Locutions verbales

avoir besoin

avoir confiance

avoir connaissance

avoir du mal

avoir faim

avoir le temps

avoir peur

avoir pour effet

avoir raison

avoir raison de

donner des ailes

donner un tour (de clés)

être à l'heure

être au courant

faire demi-tour

faire diversion

faire en sorte

faire illusion

faire mine

faire nuit

faire office de

faire semblant

faire signe

fermer à clé

jeter un coup d'œil

mettre en alerte

mettre KO

prendre garde

prendre le temps

prendre ses jambes à son cou

prendre soin de (+ verbe)

prendre son courage à deux mains

rouer de coups

se jeter à la poursuite

se mettre à genoux

se rendre compte

Annexe 3 : Sémantisme des prédicats

agression	ameuter ; arrêter ; assassiner ; asséner ; assommer ; attacher ; attraper ; avoir ; bâillonner ; bousculer ; capturer ; délivrer ; déshabiller ; donner ; empêcher ; enfermer ; entourer ; étrangler ; faire payer ; frapper ; habiller ; immobiliser ; lier ; lutter ; maintenir ; mettre ; mettre KO ; obliger ; passer ; placer ; planter ; plaquer ; plonger ; porter ; pousser ; prendre ; projeter ; redonner ; renverser ; rouer de coups ; saisir ; séparer ; torturer ; tuer
interaction	accompagner ; aider ; aimer ; ameuter ; apporter ; attendre ; attirer ; avoir raison de ; chercher ; contacter ; croiser ; donner ; embrasser ; empêcher ; épouser ; faire ; faire diversion ; faire passer ; faire signe ; garder ; glisser ; inviter ; léguer ; libérer ; menacer ; mettre en alerte ; obéir ; offrir ; prendre ; protéger ; remplacer ; rencontrer ; retrouver ; sauver ; se lier d'amitié ; se retrouver ; s'enlacer ; s'occuper de ; sourire ; trouver
sens	apercevoir ; découvrir ; distinguer ; entendre ; épier ; jeter ; jeter un coup d'œil ; observer ; reconnaitre ; regarder ; repérer ; respirer ; se regarder ; sentir ; surprendre ; surveiller ; voir
déplacement	accélérer ; aller ; aller chercher ; amener ; apparaitre ; apporter ; approcher ; arpenter ; arriver ; atteindre ; avancer ; bondir ; commencer ; continuer ; courir ; déambuler ; dépasser ; descendre ; embarquer ; emmener ; emprunter ; entrer ; escalader ; escorter ; être sorti ; faire ; faire demi-tour ; foncer ; franchir ; grimper ; longer ; marcher ; mettre ; parcourir ; partir ; parvenir ; passant ; passer ; passer incognito ; pénétrer ; porter ; prendre ; prendre ses jambes à son cou ; quitter ; ramer ; ramper ; rattraper ; reculer ; redescendre ; rejoindre ; remettre ; remonter ; rentrer ; reprendre ; rester ; rester immobile ; retourner ; retrouver ; revenir ; s'approcher ; s'arrêter ; sauter ; s'avancer ; se cacher ; se dépêcher ; se diriger ; se disperser ; se faufiler ; se glisser ; se heurter ; se jeter ; se placer ; se précipiter ; se présenter ; se rapprocher ; se rendre ; se retrouver ; se ruer ; se sauver ; s'échapper ; s'élancer ; s'en aller ; s'enfoncer ; s'enfuir ; s'engager ; s'engouffrer ; s'évader ; s'installer ; sortir ; suivre ; tirer ; tomber ; tourner ; traverser ; venir
changement de posture	élancer ; s'affaisser ; s'assoir ; se baisser ; se coucher ; se lever ; se mettre à genou ; se pencher ; se relever ; se retourner ; s'écrouler ; s'effondrer ; tomber ; vaciller
objet	abandonner ; abattre ; acheter ; actionner ; attraper ; braquer ; cacher ; chercher ; claquer ; creuser ; déchirer ; déposer ; dérober ; déverrouiller ; dévisser ; dissimuler ; donner un tour ; échanger ; empoigner ; enfiler ; enfoncer ; entourer ; envoyer ; fermer ; filer ; garder ; jeter ; laisser ; lancer ; manger ; mettre ; ôter ; ouvrir ; porter ; pousser ; prélever ; prendre ; quitter ; racheter ; ranger ; récupérer ; refermer ; repousser ; reprendre ; resserrer ; retirer ; revêtir ; saisir ; se saisir ; s'emparer ; serrer ; sortir ; subtiliser ; tailler ; tenir ; tirer ; tomber ; tourner ; trouver ; verrouiller ; vêtir ; voler
action - divers	abandonner ; accomplir ; agir ; arriver ; atteindre ; attendre ; battre ; chercher ; commencer ; commettre ; couper ; danser ; déchirer ; découvrir ; dégager ; diffuser ; diner ; disparaitre ; donner ; dormir ; échouer ; éclairer ; empêcher ; en avoir ; enlever ; envahir ; épargner ; étreindre ; étudier ; expirer ; faillir ; faire ; faire en sorte que ; finir ; fonctionner ; frapper ; gagner ; garder ; indiquer ; jeter ; laisser ; marquer ; masquer ; mener ; mesurer ; mettre ; mourir ; naitre ; ôter ; ouvrir ; parvenir ; passer ; perdre ; pleurer ; plonger ; pousser ; prendre ; prendre son courage à deux mains ; préparer ; préserver ; profiter ; promettre ; prouver ; purger ; ranimer ; rassurer ; rattraper ; recevoir ; recommencer ; refaire ; relâcher ;

	reprendre ; résister ; respirer ; retentir ; retrouver ; réussir ; revenir ; s'abattre ; s'agiter ; s'arracher ; s'arrêter ; se cacher ; se changer ; se déclencher ; se défendre ; se déshabiller ; se dessiner ; se donner ; se lever ; se monter ; se présenter ; se rapprocher ; se réguler ; se remplir ; se reposer ; se réveiller ; se serrer ; se servir ; se suspendre ; se tatouer ; s'échapper ; s'effondrer ; s'emballer ; s'emparer ; s'endormir ; s'entrainer ; s'entrouvrir ; serrer ; s'habiller ; somnoler ; sonner ; sortir ; souffrir ; soutenir ; s'ouvrir ; suivre ; tendre ; tirer ; tomber ; toucher ; tourner ; travailler ; trouver ; venir ; vivre ; y arriver
parole	alerter ; appeler ; avertir ; avouer ; chuchoter ; converser ; crier ; demander ; dire ; hurler ; interpeller ; intimer ; murmurer ; ordonner ; prévenir ; raconter ; répondre ; rire ; saluer ; se répéter ; se taire ; signaler
cognition	apprendre ; avoir ; avoir connaissance ; calculer ; comprendre ; compter sur ; connaitre ; croire ; décider ; douter ; élaborer ; espérer ; être en train de rêver ; faire ; garder ; hésiter ; imaginer ; oser ; oublier ; penser ; perdre ; planifier ; prendre garde ; prendre son courage à deux mains ; prévoir ; prévu ; prier ; réfléchir ; repenser ; reprendre ses esprits ; rêver ; s'accorder ; s'apercevoir ; s'assurer ; savoir ; se demander ; se rappeler ; se remémorer ; se rendre compte ; se savoir condamné ; se souvenir ; s'échafauder ; s'en ficher ; s'habituer ; trouver ; veiller
état	aller ; armé ; assis ; assuré ; avoir ; avoir confiance ; avoir de la chance ; avoir du bon ; avoir été heureux ; avoir faim ; avoir le droit ; avoir peur ; avoir pour effet ; avoir raison ; baigner ; cassé ; c'est ; déboucher ; demander ; désigner ; devenir ; devenir fou ; donner ; doté ; durer ; éclairer ; en arriver ; en être autrement ; enfermé ; enfoncé ; être ; être à la rêverie ; être à l'heure ; être abandonné ; être absorbé ; être affiché ; être amoureux ; être arrivé ; être attaché ; être au courant ; être banal ; être bandée ; être beau ; être bon ; être certain ; être confiant ; être dangereux ; être éclairé ; être effrayé ; être en état d'alerte ; être en évidence ; être encerclé ; être enfermé ; être enfoncé ; être entreposé ; être étendu ; être facile ; être fasciné ; être heureux ; être inconnu ; être inconscient ; être intelligent ; être là ; être le gardien ; être le meilleur ; être le moment de vérité ; être le moyen de ; être le nom ; être le seul gardien ; être libre ; être loin ; être long ; être malade ; être mort de peur ; être optimal ; être ouvert ; être ouverte ; être parfait ; être posté ; être pressé ; être prêt ; être proche ; être prudent ; être rayé ; être recouvert ; être réel ; être repérable ; être sa cellule ; être silencieux ; être son prisonnier ; être sorti ; être suffisant ; être sûr ; être surpris ; être triste ; être trop tard ; être un homme libre ; être un jeune fermier ; être un moyen de ; être un prince ; être un rêve ; être une simulation ; être vide ; faire illusion ; faire mine ; faire office ; habiter ; heureux ; il est temps ; il fait ; il faut ; il reste ; il s'agit ; il tient ; il y a ; infester ; mener ; mériter ; ouvert ; paraitre ; paraitre interminable ; passer ; plaquer ; poser ; posséder ; posté ; présent ; pris ; regroupé ; ressembler ; rester ; rester interloqué ; rester silencieux ; se faire passer ; se passer ; se sentir ; se situer ; se trouver ; sembler ; sembler être plus léger ; s'imposer ; simuler ; suspendre ; tendu ; trembler ; vivre
pseudo-état	attendre ; baigner ; déranger ; dissimuler ; donner des ailes ; éblouir ; éclairer ; enjamber ; étouffer ; manquer ; observer ; pendre ; protéger ; provoquer ; recouvrir ; saisir ; s'annoncer ; séparer ; tenir

Annexe 4 : Types de lieu (corporel et externe)

Corporel	à la ceinture ; à la ceinture de l'uniforme ; à la figure du gardien ; à la gorge du gardien ; à la main ; à la tête ; à l'arrière de la tête ; à l'intérieur de sa veste ; à l'oreille ; à sa ceinture ; au cou du gardien ; au fond de lui ; au niveau de la bouche ; au niveau de sa poitrine ; au niveau des premières phalanges ; au visage ; au visage du gardien ; autour de la taille du gardien ; autour de sa prise ; autour de son poignet ; aux joues ; contre son cœur ; dans la main du fonctionnaire ; dans la poche ; dans le bras ; dans le regard ; dans le ventre ; dans le visage ; dans les yeux ; dans leur esprit ; dans sa chaussette ; dans sa misère ; dans ses bras ; dans son pantalon ; de lui ; de sa poche ; de ses pensées ; de son visage ; du corps inerte ; en lui ; sous la gorge ; sur la bouche ; sur la gorge ; sur le corps ; sur le crâne du gardien ; sur le dos ; sur le gardien ; sur le geôlier ; sur lui ; sur sa nuque ; sur sa tête ; sur son dos ; sur son gardien ; sur son visage ; vers la tête du gardien ; jusqu'à ses poumons
Externe	à chaque coin ; à destination ; à droite ; à gauche ; à la cafeteria ; à la cantine ; à la dernière porte ; à la droite de celui qui était désormais son prisonnier ; à la frontière ; à la porte de sécurité ; à la porte d'entrée ; à la salle où est entreposé le linge sale ; à la sortie ; à l'accueil ; à l'air libre ; à l'angle du dernier couloir ; à l'arrière du camion ; à l'autre bout de sa cellule ; à l'autre extrémité du couloir ; à l'entrée de la prison ; à l'étage supérieur dans la salle de garde ; à l'infirmerie ; à l'intérieur ; à l'intérieur de la cellule où il était enfermé depuis plusieurs années ; à l'intérieur de l'étroit passage ; à l'opposé de là où il se trouvait réellement ; à sa poursuite ; à terre ; à travers celui-ci ; à travers champ ; à travers la serrure ; à travers la ville ; à travers le long couloir ; à travers les rues sombres et désertes du centre-ville ; à travers l'ouverture ; à une autre porte fermée et gardée par un gardien de nuit ; à une camionnette ; à une petite maison en briques rouges ; au bord de l'autoroute ; au bout de ces longs couloirs ; au câble ; au lit ; au loin ; au milieu des cadavres qu'il venait de faire ; au parloir ; au point de rendez-vous ; au sol ; au sous-sol ; au toit de la prison ; au vestiaire ; au-dessus du vide ; autour ; autour de lui ; aux barbelés ; aux toilettes ; bien loin ; chez les infirmiers ; chez son frère ; contre le mur ; contre le mur ; contre les barreaux de la cellule ; contre un mur ; dans 2 km ; dans ce conduit ; dans ce système ; dans cette ferme ; dans cette gigantesque serrure ; dans cette prison ; dans de vieux draps sales ; dans la barque ; dans la bibliothèque ; dans la boite à gant ; dans la cellule ; dans la cellule qui avait été la sienne pendant dix longs mois ; dans la cellule sombre ; dans la couchette de la cellule ; dans la cour ; dans la cour de la prison ; dans la grande salle ; dans la mer noire que la lune n'éclairait qu'à peine ; dans la montagne ; dans la porte ; dans la prison ; dans la prison la mieux gardée du pays ; dans la rue ; dans la salle la plus proche ; dans la salle où il se trouvait ; dans la serrure ; dans la serrure de la porte ; dans la tour où était enfermée sa sœur ; dans laquelle ; dans le bâtiment ; dans le camion ; dans le conduit ; dans le couloir ; dans le couloir de la prison ; dans le couloir en face ; dans le grillage ; dans le miroir ; dans le mouchoir imbibé de chloroforme ; dans le vide ; dans les catacombes du château ; dans les couloirs ; dans les environs de la prison ; dans les miradors ; dans l'escalier ; dans leur cellule ; dans l'obscurité de cette nuit sans lune ; dans l'obscurité de la nuit ; dans l'obscurité de la prison ; dans l'un des cartons ; dans sa cellule ; dans son lit ; dans tout le château ; dans un autre bâtiment ; dans un autre village perdu dans les montagnes ; dans un cagibi ; dans un carton ; dans un couloir ; dans un couloir qui le mena jusqu'à un escalier ; dans un cul-de-sac ; dans un cul-de-sac éclairé par une fenêtre grillagée ; dans un morceau de métal ; dans un village voisin ; dans une forêt ; dans une mare de sang ; dans une petite crique où un bateau les attendait ; dans une petite station essence ; dans une rue qui lui était inconnue ; de ce petit enfer ; de cet enfer ; de cette prison ; de cette prison où il était déjà depuis quatre ans ; de là ; de la cellule ; de la fenêtre à barreaux ; de la prison ; de l'autre côté du grillage ; de l'enceinte de la prison ; de plus près ; de sa cachette ; de sa cellule ; de son côté ; de son lit ; dedans ; dehors ; depuis sa cellule ; derrière la porte ; derrière les rideaux ; derrière lui ; derrière un petit

	mur ; derrière lui ; des toits ; devant ; devant d'innombrables portes de cellules ; devant le gardien ; devant l'une des cellules ; devant quatre hommes ; devant son ancienne cellule dans laquelle le véritable gardien, tétanisé, restait silencieux ; d'ici ; du camion ; du château ; du haut de leur mirador ; du pays ; en ; en Europe ; en forêt ; en prison ; en vue ; encore dans la serrure de la porte ; entre deux poteaux électriques ; entre la mer et lui ; entre les rayons lumineux des projecteurs actionnés par les gardiens, du haut de leur mirador ; hors de la prison ; hors de sa cellule ; ici ; jusqu'à la bibliothèque ; jusqu'à la grande porte d'entrée ; jusqu'à la mer ; jusqu'à l'extérieur de la prison ; jusqu'à un escalier ; jusqu'aux cuisines ; jusque dans son lit ; juste à côté de lui ; juste derrière ; juste l'étage en-dessous ; là ; la cellule que Dimitri trouvait tellement humiliante ; le long du couloir ; le sol de la cellule de Dimitri ; loin de ce pays ; loin de la guerre ; loin du gardien qui surveillait ; non loin de là ; où ; où Dimitri possédait une ferme léguée par son parrain, où ; par derrière ; par la porte qui donnait sur l'arrière-cour ; par le trou qu'il avait creusé ; par terre ; par-dessus son épaule ; parmi ces voix ; parmi toute la foule des prisonniers ; partout ; plus en avant ; pour la Crète où Dimitri possédait une fermée léguée par son parrain ; pour la ville ; dans sa cellule ; sous ce ciel étoilé ; sous le lit ; sous son matelas ; sous son oreiller ; sur la plage ; sur l'arrière-cour ; sur le bas de son drap ; sur le compteur ; sur le lit ; sur le sol ; sur le trousseau de clefs encore dans la serrure de la porte de la cellule de Dimitri ; sur sa route ; sur son chemin ; sur un rocher ; sur une grande pièce où des ombres dotées de parole conversaient ; sur une rue déserte ; tout autour ; trente mètres plus loin ; une fois dehors ; une fois sur le toit ; vers celui-ci ; vers la droite ; vers la liberté ; vers la porte ; vers la porte qui était à moitié ouverte ; vers la salle des contrôles dans laquelle le planning de chacun des gardiens était affiché ; vers la sortie ; vers la sortie fortement surveillée de la prison ; vers l'aile sud ; vers le ciel qui se remplissait silencieusement d'un immense disque métallique ; vers le gardien ; vers le passage secret ; vers le vestiaire ; vers l'entrée du bâtiment ; vers lui ; vers un avenir plus libre ; vers une autre cellule ; vers une autre rangée de cellules ; vers une voiture de police ; y

Annexe 5 : Champs lexicaux (CL) et catégories des référents

PERSONNAGE	Ami	ami ; camarade ; compagnon ; complice ; copain ; voisin
	Famille	amant ; cousin ; enfant ; famille ; femme ; fiancé ; fille ; frère ; mari ; parrain ; père ; sœur
	Contes	fée ; lutin ; prince ; princesse ; roi
	Prison	administration ; agent ; bourreau ; collègue ; détenu ; fonctionnaire ; garde ; gardien ; geôlier ; monde ; personnel ; policier ; prisonnier ; service ; soldat ; système
	Animal	chat ; rat
	Générique	Dernier ; foule ; groupe ; homme ; humain ; monde ; personne ; population ; tout le monde
	Autres	Athlète **(sport)** ; boxeur (**sport)** ; commis **(repas)** ; faible **(sentiment)** ; fermier **(divers)** ; infirmier **(infirmerie)** ; passeur **(fuite)** ; petit **(divers)**
OBJET	Corps	casquette ; ceinture ; chaussette ; habit ; montre ; pantalon ; poche ; sous-vêtements ; tenue ; uniforme ; veste ; vêtement
	Bâtiment	ampoule ; brique ; cloche ; couchette ; haut-parleur ; lampe ; linge ; lit ; matelas ; meuble ; miroir ; oreiller ; rideaux ; poussière ; projecteur ; serrure
	Fuite	bâillon ; bandeau ; chloroforme ; clé ; clef ; échelle ; fiole ; lame ; menotte ; pass ; pince ; sifflet ; talky-walky ; trousseau
	Violence	arme ; balle ; couteau ; matraque ; mitraillette ; revolver
	Trajet	barque ; bateau ; bip ; boite à gant ; camion ; camionnette ; compteur ; disque ; rame ; voiture
	Nature	branche ; étoile ; lune ; pierre ; pluie ; rayon ; soleil
	Repas	bol ; casserole ; eau ; fourchette ; gamelle ; nourriture ; pain ; pizza ; plateau ; soupe ; vaisselle
	Générique	détail ; morceau ; objet ; tissu
	Autres	carton (déménagement) ; cirage (matière) ; drap (linge) ; livre (livre, divertissement) ; marchandise (divers) ; métal (matière) ; mouchoir ; outil (infirmerie) ; pot (divers) ; produit (infirmerie)
ENVIRONNEMENT	Bâtiment	aile ; angle ; barrière ; bâtiment ; câble ; cagibi ; catacombe ; château ; coin ; conduit ; couloir ; cour ; cul-de-sac ; enceinte ; entrée ; escalier ; étage ; grille ; fenêtre ; maison ; marche ; mur ; passage ; persienne ; pièce ; porte ; poteau ; poussière ; salle ; seuil ; sol ; sortie ; sous-sol terre ; toilette ; toit ; tour
	Prison	accueil ; barbelé ; barreau ; bibliothèque ; cafeteria ; cantine ; cuisine ; cellule ; grillage ; infirmerie ; mirador ; parloir ; prison ; vestiaire ; zone
	Trajet	autoroute ; chemin ; destination ; kilomètre ; mètre ; passage ; route ; station ; voie

	Nature	arbre ; champ ; ciel ; continent ; crique ; forêt ; ile ; mer ; montagne ; pays ; pelouse ; plage ; rocher
	Ville	centre-ville ; faubourg ; village ; ville ; rue
	Localisation	bas ; bout ; côté ; direction ; droite ; environ ; extérieur ; extrémité ; gauche ; intérieur
	Générique	emplacement ; endroit ; environnement ; espace ; lieu ; place
	Autres	blanchisserie **(linge)** ; enfer **(divers)** ; ferme **(habitation)**
PARTIE DE L'HUMAIN	Corps	arrière ; avant ; bouche ; bras ; cerveau ; coeur ; corps ; cou ; crâne ; dent ; doigt ; dos ; épaule ; figure ; gorge ; jambe ; joue ; lèvre ; main ; moustache ; nez ; nuque ; oeil ; oreille ; phalange ; pied ; poignet ; poitrine ; poumon ; sang ; souffle ; tête ; ventre ; visage
	Parole	cri ; grognement ; parole ; rire ; silence
	Cognition	choix ; erreur ; hésitation ; idée ; imprévu ; rêverie ; souvenir ; pensée
	Sens	attention ; fraicheur ; inattention ; instinct ; observation ; regard ; sens ; sensation ; vue
	Sentiment	admiration ; calme ; courage ; culpabilité ; désespoir ; désir ; effort ; envie ; espoir ; gémissement ; joie ; patience ; peur ; satisfaction ; sentiment ; souffrance ; surprise
CONCEPT	Corps	crampe ; esprit ; état ; geste ; kilo ; nom ; ombre ; pas ; respiration ; rôle ; taille ; vie ; voix
	Parole	discussion ; histoire ; message ; mot ; récit ; réponse ; terme
	Cognition	notion ; rêve ; révélation ; sujet ; vérité ;
	Sens	aération ; air ; ambiance ; brillance ; bruit ; lueur ; lumière ; monde ; obscurité ; parfum ; pénombre ; son ; vide
	Sentiment	dur ; énergie ; faiblesse ; force ; intensité ; meilleur ; misère ; répit ; soupir
	Prison	arrivée ; captivité ; contrôle ; complicité ; emprisonnement ; état d'alerte ; innocence ; ; peine ; peine de mort ; pénitencier ; planning ; preuve ; relève ; sécurité ; séjour ; visite ; vol
	Violence	altercation ; bagarre ; cadavre ; chaos ; choc ; coup ; guerre ; lutte ; menace ; meurtre ; mort ; prise ; sommation ; tir ; trou ; vengeance
	Trajet	allers et venues ; demi-tour ; frontière ; issue ; ouverture ; point de RDV ; trajet ; voyage
	Fuite	alarme ; alerte ; aventure ;cachette ; course ; départ ; disparition ; enlèvement ; évasion ; faille ; fuite ; liberté ; pas ; piège ; plan ; planification ; poursuite ; recherche ; signe ; simulation ; solution ; tentative ; trou
	Divertissement	jeu ; bal ; fête ; série
	Comportement	habitude ; rangement ; tâche ; visite
	Générique	action ; moyen ; ouvrage

	Autres	aboiement **(animal) ;** agitation **(divers)** ; déménagement **(déménagement)** ; destin **(divers)** ; échange **(divers)** ; réfection **(repas)**
TEMPS	Divers	an ; année ; anniversaire ; aube ; avenir ; commencement ; début ; départ ; été ; éternité ; fin ; fois ; heure ; instant ; jour ; journée ; lendemain ; matin ; midi ; minute ; mois ; moment ; nuit ; pause ; séance ; seconde ; semaine ; soir ; temps
QUANTITE	Divers	compte ; deux ; filet ; importance ; lambeau ; rangée ; reste ; tout

Annexe 6 : Champs lexicaux des prédicats

Prison	avouer ; condamner ; être arrêté ; être capturé ; être enfermé ; enfermer ; garder ; voler
Violence	armé ; assassiner ; assommer ; attacher ; bâillonner ; bousculer ; empêcher ; étrangler ; être dangereux ; faire payer ; faire vaciller ; frapper ; immobiliser ; lier ; lutter ; maintenir ; menacer ; mettre ko ; mourir ; obliger ; planter ; plaquer ; pousser ; projeter ; résister ; renverser ; rouer de coups ; se défendre ; souffrir ; torturer ; tuer
Fuite	alerter ; cacher ; chercher ; délivrer ; déverrouiller ; dissimuler ; entrouvrir ; être libre ; être ouverte ; être repérable ; faire diversion ; faire illusion ; faire signe ; fermer ; libérer ; mettre en alerte ; ouvrir ; passer incognito ; protéger ; rattraper ; refermer ; s'échapper ; se cacher ; se sauver ; sauver ; s'enfuir ; s'évader ; simuler ; subtiliser ; verrouiller
Trajet	aller ; amener ; apporter ; approcher ; arpenter ; arriver ; atteindre ; avancer ; bondir ; continuer ; courir ; déambuler ; descendre ; embarquer ; emmener ; escalader ; escorter ; foncer ; franchir ; grimper ; longer ; marcher ; mener ; parcourir ; parvenir ; passer ; pénétrer ; quitter ; ramer ; ramper ; reculer ; redescendre ; rejoindre ; remonter ; rentrer ; rester ; rester immobile ; retourner ; retrouver ; revenir ; s'approcher ; s'arrêter ; s'avancer ; s'élancer ;s'en aller ; s'enfoncer ; s'engager ; s'engouffrer ; s'installer ; se dépêcher ; se diriger ; se disperser ; se faufiler ; se glisser ; se jeter ; se placer ; se précipiter ; se rendre ; se retrouver ; se ruer ; se suspendre ; sortir ; suivre ; tourner ; traverser ; venir
Corps	déshabiller ; dormir ; enjamber ; etre inconscient ; être malade ; être réel ; expirer ; habiller ; naitre ; respirer ; revêtir ; s'affaisser ; s'assoir ; s'écrouler ; s'effondrer ; s'endormir ; s'habiller ; se baisser ; se changer ; se coucher ; se déshabiller ; se heurter ; se lever ; se mettre à genoux ; se pencher ; se relever ; se répéter ; se reposer ; se retourner ; se réveiller ; somnoler ; sourire ; tomber ; tomber évanoui ; vêtir ; vivre
Cognition	apprendre ; avoir connaissance ; calculer ; craindre ; croire ; comprendre ; compter sur ; connaitre ; décider ; douter ; espérer ; être intelligent ; étudier ; hésiter ; imaginer ; mesurer ; oser ; oublier ; penser ; prendre garde ; planifier ; prévoir ; prier ; réfléchir ; repenser ; rêver ; s'accorder ; s'apercevoir ; s'assurer ; s'échafauder ; s'en ficher ; s'habituer ; savoir ; se décider ; se demander ; se rappeler ; se remémorer ; se rendre compte ; se souvenir ; veiller ; vouloir
Parole	ameuter ; appeler ; avertir ; chuchoter ; contacter ; converser ; crier ; demander ; dire ; être silencieux ; interpeller ; intimer ; hurler ; murmurer ; ordonner ; proposer ; raconter ; répondre ; rire ; saluer ; se taire ; signaler
Sens	apercevoir ; apparaitre ; découvrir ; distinguer ; eclairer ; entendre ; épier ; être ébloui ; observer ; reconnaitre ; regarder ; repérer ; se regarder ; sentir ; surprendre ; surveiller ; voir
Sentiment	aimer ; avoir confiance ; avoir du mal ; avoir peur ; devenir fou ; être amoureux ; etre beau ; être certain ; être confiant ; être effrayé ; être facile ; être fasciné ; être heureux ; être léger ; être mort de peur ; être pressé ; etre prêt ; être prudent ; être sûr ; être surpris ; être triste ; paraitre

	interminable ; pleurer ; préférer ; rassurer ; rester interloqué ; se sentir soulagé ; souhaiter ; trembler
Comportement	faillir
Famille	embrasser ; épouser ; léguer ; s'enlacer
Ami	se lier d'amitié
Repas	avoir faim ; diner ; manger
Divertissement	danser
Temps	durer

Annexe 7 : Phrases – résultats

	ICC modèle 0, modèle 2	**LL modèle 0**	**LL modèle 2**	**Nb de paramètres dans le modèle 2**	**χ^2**	**Pseudo R^2**	**VI**	**Effet de la VI (p-value)**	**résultat**
C_{n-1}									
Pauses intra-mot	27,54%, 28,01%	-93,578	-124,905	21	$\chi_{(18)}$=31,33, p<0,05	13,94%	**(Type de sujet)**	0,089	
							(Continuité du sujet par rapport au 1er sujet de la phrase précédente)	0,054	Maintien >changement[82]
							(Présence d'un complément de temps au niveau 1)	0,095	Présent (Pst) >Absent (Abs)
							(Présence de clauses non finies)	0,092	Pst >Abs
							% de OS (=OR +OV) au niveau 1	0,02	+[83]
							(Difficulté d'accessibilité des référents)	0,063	+
Exécution	64,80%	-647,913	/	/	Rien sig. (ns)	/			
Pause inter-mots	15,23%, 18,70%	108,44	83,989	17	χ(14)=24,45, p<0,05	11,85%	**(Présence d'un adverbe de liaison)**	0,082	Pst< Abs
							(ASL / phrase n-1)	0,053	+
							(Profondeur syntaxique maximum)	0,064	+
							(Densité lexicale)	0,058	-
Écriture interne	12,53%	-261,523	/	/	Rien sig. (ns)	/			
C_1									
Pause initiale	20,04%,	361,52	333,58	17		10,61%	**Présence d'un prédicat de 1er plan**	0,025	Pst >Abs

[82] Le signe « > » indique que la variable chronométrique (durée de la pause ou temps d'exécution) est plus longue ; les signes « > » et «< » sont utilisés pour les VI catégorielles

[83] Le signe « + » indique une corrélation positive entre la VI et la VD : par exemple, plus il y a de « % de OS pour le niveau 1 », plus la moyenne des pauses intra-mots est longues ; les signes « + » et « - » sont utilisés pour les VI continues

	19,66%				χ(14)=13,84, ns		**Fréquence des prédicats**	0,032	-
							(% de syntagmes FC au niveau 1)	0,052	-
Pauses intra – mots	31,85%, 35,85%	-73,254	-115,2	27	χ(24)=41,95, p<0,02	17,14%	**Présence d'un adverbe de liaison**	0,04	Pst >Abs
							Type de sujet	0,02	Ø>autres
							Fréquence des prédicats	0,025	-
							(% de syntagmes FC au niveau 1)	0,071	-
							Fréquence lexèmes – Manulex	0,034	-
Exécution	39,11%, 43,35%	-438,94	-463,667	16	χ(13)=24,73, p<0,05	11,17%	**% de OS (=OR +OV) au niveau 1**	0,018	+
							(Profondeur syntaxique maximum)	0,093	+
							(Fréquence lexèmes – Manulex)	0,084	-
Pause inter-mots	18,52%, 21,64%	95,145	49,938	27	χ(24)=45,21, p<0,01	18,09%	**Présence d'un prédicat de 1er plan**	0,007	Pst >Abs
							Présence d'un complément de Lieu au niveau 1	0,032	Pst >Abs
							(ASL / texte précédent)	0,096	+
							(% de syntagmes FC au niveau 1)	0,062	+
							(% de OS (=OR +OV) au niveau 1)	0,061	+
							Profondeur syntaxique maximum	0,047	+
							Fréquence lexèmes – Manulex	0,018	-
Écriture interne	8,84%, 12,40%	-232,199	-289,898	27	χ(24)=57,70, p<0,001	22,65%	**(Présence d'un adverbe de liaison)**	0,062	Pst >Abs
							Type de sujet	0,019	Ø>personnage non Dimitri
							Profondeur syntaxique maximum	0,017	+
							(Fréquence lexèmes – corpus)	0,097	-
							Fréquence lexèmes – Manulex	0,006	-

Annexe 8 : Clauses matrices – résultats

	ICC modèle 0, modèle 2	**LL modèle 0**	**LL modèle 2**	**Nb de paramètres dans le modèle 2**	**χ^2**	**Pseudo R^2**	**VI**	**Effet de la VI (p-value)**	**Résultat**
C_{n-1}									
Pause initiale	20.50%, 21.18%	231.951	169.296	37	$\chi(34)=62.655$, $p<0.001$	38.99%			
							Présence de syntagme FC	,006	pst > abs
							Type de complément de temps	,001	balise< aucun, autre
							Temporalité du prédicat	,044	simultané< général
							(Sémantique du prédicat)	,059	ns
							Type de sujet	,007	autre personnage >Dimitri, non personnage
							Difficulté d'accessibilité des référents	,031	-
							(Profondeur syntaxique maximum)	,064	-
							% de référents avec insertion de l'item lexical	,005	-
							% de référents avec réinsertion de l'item lexical	,027	-
							% de référents avec reprise de l'item lexical	,011	-
							Densité lexicale	,001	-
							Fréquence des lexèmes – corpus	,038	+
							% de référent « personnage »	,004	-
Pauses intra –mots	18.63%, 24.59%	-12.844	-34.575	15	$\chi(12)=-21.731$, $p<0.05$	19.00%			
							Continuité du sujet par rapport à la clause précédente	,005	maintien >changement

							(Nature grammaticale du sujet)	,056	(syntagme nominal > ellipse)
							Nombre de séquences	,021	-
Exécution	60.60%, 67.49%	-282.067	-321.076	29	$\chi(26)=39.009$, $p<0.05$	30.52%			
							Présence d'une liaison	,003	pst< abs
							Temporalité du prédicat	,044	(général > passé, passé dans histoire)
							Nature grammaticale du sujet	,000	ellipse > nom propre, pronom
							% de référents avec réinsertion de l'item lexical	0.021	+
							Densité lexicale	0.022	-
							Fréquence des lemmes – Manulex	,005	-
Pause inter-mots	23.69%, 27.93%	28.934	-20.536	34	$\chi(31)=49.47$, $p<0.05$	33.96%			
							(Présence de complément de lieu)	,068	pst< abs
							Continuité du sujet par rapport à la clause précédente	,019	maintien > changement
							Fréquence des lemmes – corpus	0.02	+
							Fréquence des lemmes – Manulex	0.011	-
Écriture interne	20.82%, 21.88%	-143.63	-160.724	12	$\chi(9)=17.094$, $p<0.05$	13.13%			
							Type de complément de temps	,045	autre > balise
							Densité lexicale	,048	-
C_n									
Pause initiale	16.48%, 23.32%	216.829	149.724	37	$\chi(34)=67.105$, $p<0.001$	43.55%			
							(Présence de complément de lieu)	0,058	pst > abs
							Présence d'une liaison	0	pst > abs
							Profondeur syntaxique maximum	0,029	-
							(Densité lexicale)	0,063	+
							Fréquence des lexèmes –Manulex	0,049	+

Pauses intra –mots	20.71%, 30.50%	-2.822	-45.18	27	χ(24)=42.358, p<0.02	32.78%		
							Sémantique du prédicat	0.006	état > déplacement
							Continuité du sujet par rapport à la clause précédente	0.012	maintien < changement
							Nature grammaticale du sujet	0.004	ellipse< nom propre, pronom
							Profondeur syntaxique maximum	0.04	+
							% de référents avec réinsertion de l'item lexical	0.016	-
							Nombre de séquences	0.032	-
Exécution	81.95%, 42.65%	-397.258	-446.994	37	χ(34)=49.736, p<0.05	39.21%		
							Présence de complément de lieu	0,024	pst > abs
							(Présence d'une liaison)	0,055	pst< abs
							Plan du prédicat	0,017	1er > arrière
							Type de sujet	0,01	Dimitri > autre personnage
							Continuité du sujet par rapport à la clause précédente	0,005	maintien< changement
							Présence d'un verbe modal	0,037	pst > abs
							(% de référents avec reprise de l'item lexical)	0,062	-
							(Fréquence des lemmes – corpus)	0,058	-
Pause inter-mots	12.74%, 14.66%	78.638	12.707	37	χ(34)=65.931, p<0.001	41.10%		
							Présence de syntagme FC	0,046	pst > abs
							Type de complément de temps	0,047	aucun > balise
							Type de sujet	0,033	(Dimitri > autre personnage)
							% de référents avec réinsertion de l'item lexical	0,018	-
							% de référents avec reprise de l'item lexical	0,008	-

							Densité lexicale	0,031	+
Écriture interne	20.49%, 54.30%	-121.534	-217.552	37	χ(34)=96.018, p<0.001	62.49%			
							Présence de syntagme FC	,005	pst > abs
							(Présence de complément de lieu)	,057	pst > abs
							Type de complément de temps	,039	aucun > balise
							(Sémantique du prédicat)	,066	état > déplacement
							Type de sujet	,000	Dimitri > autre personnage, non personnage
							Nature grammaticale du sujet	,030	ns
							(Profondeur syntaxique maximum)	,096	+
							% de référents avec insertion de l'item lexical	,040	+
							Nombre de séquences	,005	-

Annexe 9 : Clauses subordonnées – résultats

	ICC modèle 0, modèle 2	**LL modèle 0**	**LL modèle 2**	**Nb de paramètres dans le modèle 2**	**χ^2**	**Pseudo R^2**	**VI**	**Effet de la VI (p-value)**	**résultat**
C_{n-1}									
Pause initiale	10.70%, 2.35%	171.282	135.162	26	$\chi(23)=36.12$, $p<0.05$	24.08%			
							(Présence d'un complément de temps)	,051	pst < abs
							Nature grammaticale du sujet	,014	pronom > syntagme nominal
							% de référents avec reprise de l'item lexical	,006	-
Pauses intra –mots	18.00%, 14.73%	1.25	-32.285	24	$\chi(21)=33.535$, $p<0.05$	25.76%			
							Présence de syntagme FC	,020	pst < abs
							Présence d'un complément de lieu	,002	pst > abs
							Plan du prédicat	,037	1er > arrière
							Sémantique du prédicat	0,001	déplacement, état < autres
							Présence d'un verbe modal	0,006	pst > abs
							(Nombre de séquences)	0,072	+
							Nombre de référents	0,005	-
Exécution	80.68%, 91.12%	-302.432	-360.116	33	$\chi(30)=57.684$, $p<0.005$	53.95%			
							Sémantique du prédicat	0	action < autres
							Type de sujet	0,031	non personnage < Dimitri, autre personnage
							(Présence d'un verbe modal)	0,098	Pst > Abs
							Fréquence du prédicat	0,022	-
							% de référents avec reprise de l'item lexical	0,032	-

Pause inter-mots	14.98%, 15.67%	83.693	45.9	28	$\chi(25)=37.793$, $p<0.05$	30.83%			
							(Nature grammaticale du sujet)	,082	(pronom > syntagme nominal)
							(Profondeur syntaxique maximum)	,090	+
Écriture interne	10.38%, 3.87%	-79.714	-95.748	11	$\chi(8)=16.034$, $p<0.05$	9.42%			
							(Nature grammaticale du sujet)	,070	(pronom > syntagme nominal)
							(Profondeur syntaxique maximum)	,060	+
							Nombre de référents	,044	-
							% de référent « personnage »	,049	-
C_n									
Pause initiale	3.81%, 11.53%	151.533	100.711	33	$\chi(30)=50.822$, $p<0.02$	42.34%			
							Sémantique du prédicat	0,01	action < déplacement, état
							(Difficulté d'accessibilité des référents)	0,081	+
							(% de référents avec réinsertion de *l'item lexical)*	0,093	-
							Nombre de séquences	0,028	+
							(% de référent « personnage »)	0,078	+
Pauses intra –mots	27.62%, 33.55%	-21.937	-56.35	25	$\chi(22)=34.413$, $p<0.05$	31.12%			
							Présence de syntagme FC	,005	pst < abs
							Temporalité du prédicat	,044	ns
							(Nature grammaticale du sujet)	,063	(syntagme nominal > ellipse)
							(Fréquence du prédicat)	,076	+
							(% de référents avec insertion)	,071	-
							% de référents avec réinsertion de *l'item lexical*	,050	-

Exécution	83.97%, 90.62%	-343.36	-387.158	33	$\chi(30)=43.798$, $p<0.05$	43.56%			
							(Continuité du sujet par rapport à la clause précédente)	0,067	maintien > changement
							Nature grammaticale du sujet	0,05	syntagme nominal > pronom
							(Profondeur syntaxique maximum)	0,095	+
Pause inter-mots	21.75%, 17.24%	19.333	-56.115	33	$\chi(30)=75.448$, $p<0.001$	49.96%			
							Présence d'un complément de lieu	0,016	pst > abs
							Plan du prédicat	0,032	1er < arrière
							Temporalité du prédicat	0,016	ns
							Nombre de séquences	0,015	+
Écriture interne	23.53%, 31.00%	-121.025	-183.499	33	$\chi(30)=62.474$, $p<0.001$	48.00%			
							Temporalité du prédicat	0,031	passé > irréel
							(Nature grammaticale du sujet)	0,065	
							Densité lexicale	0,012	+

Annexe 10 : Clauses autonomes – résultats

	ICC modèle 0, modèle 2	**LL modèle 0**	**LL modèle 2**	**Nb de paramètres dans le modèle 2**	**χ^2**	**Pseudo R^2**	**VI**	**Effet de la VI (p-value)**	**résultat**
C_{n-1}									
Pause initiale	7.99%, 9.66%	560.146	472.731	36	$\chi(33)=87.415$, $p<0.001$	19.18%			
							(Présence d'un adverbe de liaison)	0,068	pst > abs
							Fréquence des prédicats	0,007	-
							(Nombre de séquences)	0,095	-
Pauses intra – mots	27.22%, 32.73%	-37.58	-63.117	18	$\chi(15)=25.537$, $p<0.05$	10.58%			
							(Type de complément de lieu)	,078	ns
							Nature grammaticale du sujet	,012	ellipse > pronom
Exécution	45.65%	-586.853	/	/	ns	/			
Pause inter-mots	13.71%	156.713	/	/	ns	/			
Écriture interne	15.82%, 15.64%	-277.461	-270.611	17	$\chi(14)=23.745$, $p<0.05$	8.96%			
							(Présence d'un adverbe de liaison)	,063	pst > abs
C_n									
Pause initiale	15.07%, 20.23%	490.9	409.882	36	$\chi(33)=81.018$, $p<0.001$	26.69%			
							Présence de syntagmes FC	0,009	pst < abs
							Présence d'un adverbe de liaison	0,044	pst < abs
							Plan du prédicat	0,011	1er > arrière
							(Fréquence des lexèmes – corpus)	0,081	-
Pauses intra – mots	31.81%, 34.80%	-80.848	-116.463	26	$\chi(23)=35.615$, $p<0.05$	12.93%			
							(Temporalité du prédicat)	,086	ns

							(Nombre de séquences)	,055	-
Exécution	37.36%, 40.04%	-534.198	-570.319	26	χ(23)=36.121, p<0.05	12.85%			
							(Type de complément de temps)	,091	
							Plan du prédicat	,012	1er > arrière
							(Difficulté d'accessibilité des référents)	,066	+
							% de référents avec reprise de *l'item lexical*	,041	+
							Fréquence des lexèmes – Manulex	,008	-
							Fréquence des lexèmes – corpus	,043	+
Pause inter-mots	22.70%, 24.14%	171.032	107.618	36	χ(33)=63.414, p<0.05	20.62%			
							Sémantique du prédicat	0,05	(état > action)
							(Type de sujet)	0,068	(Dimitri > autre personnage)
							(Continuité du sujet par rapport au 1er sujet de la phrase précédente)	0,074	maintien < changement
							Nature grammaticale du sujet	0,031	(ellipse > nom propre, pronom)
							(Profondeur syntaxique maximum)	0,077	+
Écriture interne	12.65%, 13.48%	-230.952	-299.812	36	χ(33)=68.86, p<0.02	21.93%			
							Sémantique du prédicat	,037	état < cognition, parole, sens ; déplacement
							(% de référents avec insertion de *l'item lexical)*	,060	+
							% de référents avec reprise de *l'item lexical*	,044	+
							Densité lexicale	,039	+
							Fréquence des lemmes – Manulex	,000	-