
Segments répétés, jets textuels et autres routines. Quel niveau de pré-construction ?

Repeated segments, bursts of writing and other routines: What level of pre-construction?

Georgeta Cislaru et Thierry Olive



Édition électronique

URL : <http://journals.openedition.org/corpus/2846>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 15 janvier 2017

ISBN : 1638-9808

ISSN : 1638-9808

Référence électronique

Georgeta Cislaru et Thierry Olive, « Segments répétés, jets textuels et autres routines. Quel niveau de pré-construction ? », *Corpus* [En ligne], 17 | 2017, mis en ligne le 15 janvier 2018, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/corpus/2846>

Ce document a été généré automatiquement le 8 septembre 2020.

© Tous droits réservés

Segments répétés, jets textuels et autres routines. Quel niveau de pré-construction ?

Repeated segments, bursts of writing and other routines: What level of pre-construction?

Georgeta Cislaru et Thierry Olive

- 1 Comme le souligne Arrivé (1998), un des premiers problèmes auquel se confronte l'analyse linguistique est celui de la délimitation des unités et de leur degré de complexité analytique. Dans cet article, nous considérons le segment comme une unité d'analyse linguistique. Bien qu'étant avant tout une catégorie opératoire, il permet d'aborder de manière structurée et cohérente des ensembles plus complexes de productions langagières (textes, discours, grands corpus, etc.) ou de développer des analyses contextualisées de phénomènes lexicaux, sémantiques, syntaxiques, etc. De ce fait, le segment tend à accéder au statut d'unité de langage¹, grâce à ses propriétés syntaxiques, sémantiques ou uniquement statistiques².
- 2 Nous nous proposons dans cet article d'apporter un point de vue en production, en analysant les segments identifiés au niveau du processus de production écrite, c'est-à-dire les unités langagières produites de manière spontanée entre deux pauses. Ces unités, que nous appelons jets textuels, ont été enregistrées grâce à l'outil de suivi de rédaction *Inputlog*. Nous avons tenté de caractériser les jets textuels en les mettant en regard avec divers types d'unités segmentales de langage, qu'elles soient issues de la segmentation des textes finalisés, ou reconnues comme des segments préfabriqués propres à un système lexical ou syntaxique, c'est-à-dire des figements au sens large du terme. Nous nous sommes plus particulièrement penchés sur un parcours de segmentation parallèle, en production (jets textuels) et en réception (segments répétés), pour interroger quelques concepts théoriques et méthodologiques qui sous-

tendent les démarches de segmentation : pré-construit, automatisme, mémorisation, figement, enracinement cognitif, fréquence, répétition :

- Comment prendre en compte la fréquence dans la définition et la délimitation des segments ? Qu'est-ce qui est récurrent, qu'est-ce qui se répète ? Les formes graphiques ? Les mots ? Les patrons syntaxiques ? Les champs conceptuels des lexèmes ? Les unités sémantiques des lexèmes intervenant dans le même patron syntaxique ?
 - Un segment relève-t-il nécessairement d'une pré-construction ? La préconstruction, le préfabriqué, le prêt-à-l'emploi étant des précurseurs du figement (Gülich et Krafft 1997), comment segments et figement s'articulent-ils ?
- 3 En lien avec ces critères définitoires, on retient deux sujets de réflexion concernant les segments : leur fonctionnement textuel et leur intervention dans la cohérence sémantique.
- 4 La première section de l'article expose les principes et enjeux de différents types de segmentation à l'écrit et définit la notion de jet textuel en interrogeant son statut linguistique et textuel. La deuxième partie de l'article aborde le rôle sémantique et textuel des segments par le prisme de deux traits définitoires, la répétition et la pré-construction, et des concepts apparentés.

1. Segments et segmentation : enjeux méthodologiques et empiriques d'une unité d'analyse linguistique

- 5 Nous distinguons ici deux groupes de segments : ceux issus de la segmentation des textes, dans le but d'étudier ces derniers, et ceux identifiés en tant que formats relatifs au système, même si leur actualisation peut être soumise à des contraintes textuelles et génériques. Il s'agit enfin de situer les jets textuels par rapport à ces catégories.

1.1 Principes de segmentation textuelle : entre grammaire et statistique

- 6 La segmentation d'un texte en unités plus ou moins figées s'appuie sur les outils textométriques appliqués au corpus. La segmentation de la chaîne textuelle en unités distinctes est un préalable à toute statistique textuelle (Salem 1987 : 16). Le texte se définit dans ce cadre comme une suite finie de caractères (Salem 1987 : 33). Les unités d'analyse sont issues d'une segmentation automatique, basée sur critères graphiques appuyés par des délimiteurs spécifiques :
- L'introduction des délimiteurs de séquence permet de ne considérer comme segment et, éventuellement par la suite, comme segment répété que des suites de formes entre lesquelles ne s'intercale aucun caractère de ponctuation faible ou forte. (Salem 1987 : 79).
- 7 Cette démarche prend donc comme critères définitoires l'identité formelle et la répétition au sein même du corpus étudié : c'est notamment le cas des segments répétés (Lafon & Salem 1983), des collocations, des paquets lexicaux (ou *lexical bundles*, Biber 2009 ; Biber *et al.* 2004), etc., et, dans un esprit plus variationniste, des motifs (Longrée & Mellet 2013). Les segments ainsi obtenus sont *a priori* validés uniquement

vis-à-vis du corpus d'observation, même si d'autres corpus peuvent être mis à contribution.

- 8 L'autre démarche, bien que travaillant de plus en plus à partir de données émergent d'un corpus, s'appuie sur une grammaire pré-existante et comprend le critère du figement formel, tandis que les critères de répétition et de fréquence restent aléatoires car pas toujours réalisés au sein d'un texte ou d'un corpus ; on a recours à une fréquence abstraite, par opposition à la fréquence mesurée exploitée lors des segmentations outillées (cf. Loiseau 2015). On mentionnera ici la phraséologie au sens large (cf. Legallois & Tutin 2013), les colligations (Hoey 2005), les constructions (Goldberg 2006), etc.
- 9 Les deux démarches ne sont pas exclusives, au sens où les segments issus de l'une peuvent parfois trouver leur place dans l'autre, et vice-versa (par exemple, les motifs sont exploités pour identifier des types de textes), tandis que la notion de fréquence peut se constituer en outil explicatif (Bybee 2010) y compris pour les processus de figement ou d'enracinement cognitif (Goldberg 2006).
- 10 Le segment se définit donc soit par des frontières délimitées statistiquement, sur la base des récurrences de formes, soit par une unité de sens-forme qui est reproduite à l'identique ou acceptant des variations minimales dans les limites d'un schéma fixe. Le lien entre les deux catégories se cristallise autour de présupposés théoriques et méthodologiques où l'idée de répétition intervient à des niveaux différents : elle est définitoire pour la première catégorie de segments, et constitue une condition de stabilisation pour les segments de la seconde catégorie.
- 11 En réfléchissant au texte de l'appel, nous avons été interpellés par la définition de l'objet *segment* en tant que structure lexico-grammaticale plus ou moins figée, pré-formée, qui combine deux ou plusieurs unités. Cette définition, qui semble faire l'objet d'un consensus tacite, ne s'applique pas moins de manière parcellaire aux objets assemblés sous l'étiquette de *segments*. Ainsi, si l'on définit les préfabriqués ou les pré-construits comme des unités non compositionnelles au niveau de la formation, mais possiblement compositionnelles au niveau du sens, on met à l'écart des unités comme les segments répétés, qui sont en quelque sorte pré-construits par la textualité, mais dont la formation ne répond pas au critère de non-compositionnalité, validé uniquement au niveau du relevé automatisé en tant que chaînes récurrentes. La question qui se pose est de savoir ce qui est pré-formé, et à quel niveau. G. Schmale propose la définition suivante :

Est considérée comme préformée une unité de construction verbale et/ou non verbale, obligatoirement caractérisée par son caractère polyfactoriel et un certain degré de stabilité qui permet à un membre de la communauté langagière concernée de la reconnaître et de la réutiliser, sachant qu'elle est plus ou moins variable pour ce qui est de sa taille, sa stabilité, sa compositionnalité sémantique, sa dissémination et sa durabilité, sa saturation lexicale, son imbrication dans une situation de communication spécifique, la présence d'activités non verbales. (Schmale 2013 : 41)
- 12 Selon Wray et Perkins (2000 : 1), les unités pré-formées, ou préfabriqués, sont stockées et réactualisées telles quelles : « [they are] stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar ». Cette définition va dans le même sens que la proposition de Grunig (1997) de considérer que le figement est avant tout une question d'inscription durable dans la mémoire³. En revanche, selon Edmonds (2013 : 133), il n'y a pas de preuve formelle

d'une inscription mémorielle des phénomènes phraséologiques et l'automatisation en réception ou en production serait alors indépendante de la mémorisation. Mais que sait-on au juste au sujet des conditions de formation ou d'actualisation, compte tenu du nombre limité d'études linguistiques portant sur le processus de production langagière à l'écrit d'une part, et les variations combinatoires d'autre part (Anscombe 2011 ; voir aussi les erreurs d'actualisation telles les greffes collocationnelles étudiées par Polguère 2007) ? Il convient dès lors de s'interroger sur les cadres de pré-construction : qu'est-ce qui est pré-construit de fait, et pour quelle étendue de validité théorique et empirique ? Quels sont les mécanismes de pré-construction à l'œuvre ?

- 13 Les réponses à ces questions seront soumises à l'épreuve d'une troisième démarche de segmentation textuelle, spontanée et en amont, la segmentation en jets de production textuelle lors du processus d'écriture. Permettant d'aborder des questions liées aux automatismes cognitifs et linguistiques, la segmentation spontanée inverse quelque part la problématique, et apporte ainsi un éclairage nouveau aux questions soulevées plus haut.

1.2 Les jets textuels, des segments produits spontanément dans le processus d'écriture

- 14 Nous travaillons avec un corpus de textes dont nous observons l'écriture en temps réel, grâce au logiciel de suivi de rédaction *Inputlog* (Leijten & Van Waes 2013), qui enregistre la chronologie du processus de rédaction et de révision du texte. Le corpus d'analyse est constitué de dix dossiers d'écriture en temps réel de rapports éducatifs de la protection de l'enfance⁴. Un dossier d'écriture comprend l'enregistrement en temps réel par *Inputlog* du processus de rédaction (évolution progressive du texte et sa chronologie, pauses, opérations de révision : suppression, ajout, déplacement, remplacement), ainsi que les versions du texte arrêtées lors de chaque session d'écriture, soit entre deux et seize versions de texte avant la version définitive. Lors de la rédaction, les scripteurs alternent des moments de pause, sans écriture, avec des moments de transcription continue du texte. Les pauses se produisent parce que les rédacteurs n'ont plus d'information pour continuer leur texte, ou parce qu'ils doivent l'évaluer. Les périodes de transcription correspondent quant à elles à des moments de transcription ininterrompue de séquences textuelles, que nous appellerons « jets textuels » (le terme original anglais est « burst », Chenoweth & Hayes 2001). Ces jets textuels peuvent prendre la forme d'une lettre, d'un mot, ou d'une séquence de mots. Par exemple, l'énoncé *une cousine qui peut venir partager du temps avec elle pendant le week-end* peut être produit sous la forme suivante :

[pause] *une* [pause] *cousine* [pause] *qui* [pause] *peut* [pause] *venir* [pause] *partager* [pause] *du* [pause] *temps* [pause] *avec* [pause] *elle* [pause] *pendant* [pause] *le* [pause] *w* [pause] *EEK* [pause] – [pause] *end.* [pause]

- 15 Par définition, un jet textuel correspond donc à une séquence de texte produite lors d'un moment de transcription fluide séparé par deux périodes de préparation mentale du texte, c'est-à-dire par des pauses. Plusieurs contraintes liées au fonctionnement cognitif déterminent en partie les propriétés des jets textuels. Ainsi, de nombreuses pauses surviennent seulement en raison de contraintes mécaniques liées à l'écriture ou à la frappe au clavier (barre du t, retour à la ligne, pause entre chaque frappe) et ne sont donc pas le résultat de contraintes ou traitements cognitifs. En conséquence, il est nécessaire de distinguer un seuil en deçà duquel les pauses ne doivent pas être prises

en compte. À ce jour, l'analyse des jets textuels repose généralement sur des pauses supérieures à 2 secondes, mais d'autres durées peuvent être utilisées (voir Chenu *et al.* 2014).

- 16 D'un point de vue psycholinguistique et cognitif, la segmentation spontanée est soumise à des contraintes contextuelles (de processus, de linéarité, de situation de production) qui font valoir les possibles du code linguistique au niveau de la performance langagière. Par exemple, comme l'ont montré Chenoweth et Hayes (2001), les rédacteurs produisent des jets textuels plus longs (en durée et en mots) lorsqu'ils rédigent dans leur langue maternelle plutôt que dans une langue seconde. De même, des rédacteurs plus avancés dans l'apprentissage d'une langue seconde produisent des jets plus longs. Enfin, Connelly *et al.* (2012) ont pu constater que des enfants présentant un trouble spécifique du langage produisent des jets plus courts que des enfants typiques. Baaijen, Galbraith et De Glopper (2012) ont pour leur part fait valoir que la mise en œuvre du processus de révision est un élément déterminant des textuels : les jets textuels marquant une révision sont plus courts que les jets de production. Aucune recherche n'a, à ce jour, évalué l'impact des traitements sémantiques sur la longueur, la durée, ou la nature des jets textuels. En revanche, plusieurs travaux ont analysé l'impact des compétences de transcription des rédacteurs sur les jets. Il ressort de ces travaux que les jets textuels sont plus courts lorsque le coût cognitif de la transcription est élevé, par exemple, parce que l'outil de transcription n'est pas familier (Alves, Castro, Sousa et Strömquist 2007 ; Alves, Castro et Olive 2012) ou parce que la calligraphie n'est pas encore maîtrisée comme chez les jeunes rédacteurs (Alves, Branco et Olive 2012). De ce fait, des entraînements qui améliorent les compétences de transcription des enfants aboutissent à la production de jets textuels plus longs (Alves *et al.* 2016).
- 17 Globalement, les jets textuels évoluent donc avec la compétence des scripteurs (pour une revue, voir Olive 2014). Fonctionnellement, dans des situations de coût élevé, le rédacteur ne peut transcrire son texte et simultanément préparer le segment suivant. Il a besoin de pauses pour focaliser son attention sur la préparation du segment de texte à venir. En revanche, lorsque le coût de la production est faible, les jets textuels sont plus longs car cela permettrait au rédacteur de mettre en œuvre simultanément plusieurs processus cognitifs : il pourrait ainsi préparer le jet suivant ou réviser le texte déjà produit simultanément à la transcription. Les jets textuels seraient donc essentiellement contraints par le coût imposé par la production à la mémoire de travail du rédacteur. Cette mémoire, essentielle pour superviser et réguler les traitements cognitifs, est de capacité limitée et contraint donc fortement la mise en œuvre des traitements rédactionnels (Olive 2011 2014). On peut donc penser que les différences individuelles au niveau des capacités de mémoire de travail (et/ou à court terme) chez les rédacteurs déterminent au moins en partie les caractéristiques des jets textuels, et au minimum leur longueur et leur durée.
- 18 Par ailleurs, les jets textuels pourraient refléter des automatismes linguistiques, dans la mesure où leur production pourrait être sous-tendue par des schémas langagiers pré-construits (Olive et Cislaru 2015). Cela serait d'autant plus le cas lorsque la production est réalisée dans un cadre ou un contexte très figé, permettant au rédacteur d'utiliser des formes récurrentes car usuelles. De telles structures préfabriquées pourraient donc être récupérées en mémoire à long terme comme une seule unité. De ce point de vue, ces préfabriqués linguistiques devraient être transcrits en un seul jet textuel. Il faut

aussi noter que le recours à des formes préfabriquées a également pour conséquence de minimiser le coût des traitements cognitifs de génération du texte en les réduisant à de la récupération en mémoire à long terme. En somme, de multiples contraintes cognitives, situationnelles ou contextuelles, peuvent affecter les caractéristiques des jets textuels. Dans le cadre cet article, nous nous focalisons sur les contraintes linguistiques et discursives.

1.3 Le statut d'unité linguistique et textuelle des segments

- 19 La segmentation du texte finalisé relève avant tout d'une démarche topologique, qui envisage le texte comme un ensemble unitaire, comme un espace d'organisation discursive spécifique. Dans cette optique, « [l]a dénomination de 'segment' souligne bien le fait que ces unités sont obtenues en opérant des coupures dans la dimension syntagmatique d'un texte déjà existant » (Salem 1987 : 51).
- 20 Or les jets textuels sont les unités d'un texte en construction⁵, dont la segmentation est le fait des sujets eux-mêmes et non le fait d'une analyse outillée, comme c'est le cas des segments répétés, par exemple. La légitimité des données issues du temps réel de l'écriture s'inscrit ainsi dans une linguistique de l'usage (cf. Bybee 2010), qui propose un regard sur la production et non seulement sur la réception des textes, et soulève la question de la description et du statut linguistique des données ainsi recueillies.
- 21 Se pose, dans un premier temps, la question des conditions d'analyse lexico-sémantique ou syntaxique des segments. Alors que les segments idiomatiques font le plus souvent l'objet de descriptions sémantiques (cf. Tamba 2011), l'annotation morphosyntaxique est souvent un préalable à l'identification des colligations, des constructions et des motifs. L'approche est davantage lexicale pour ce qui est des paquets lexicaux, collocations ou segments répétés. Dans le cas des segments répétés, identifiés comme des formes (lexicales), le recours aux formes-pôle (Salem 1987 : 140 et ss), à l'observation des expansions droite et gauche des segments répétés (Salem 1987 : 52-53), aux cooccurrences et aux listes de formes pré-sélectionnées par l'analyste permet une première approche lexico-sémantique, qui est généralement couplée à des hypothèses fonctionnelles :
- [...] la dénomination *polyforme* [ex. : *sécurité sociale*] implique davantage l'idée qu'une suite de formes, dont on n'a pas encore nécessairement repéré l'existence ni a fortiori toutes les occurrences éventuelles dans le texte que l'on étudie, possède, dans ce texte sinon en langue, **une unité de fonctionnement** qui lui est propre. (Lebart et Salem 1994 : 61 ; le gras est de nous)
- 22 La description syntaxique se manifeste plutôt en creux, par l'élimination des segments répétés relevant de stéréotypes grammaticaux (*de la, à la*) et par la mise en évidence de séquences ne constituant pas des « unités en langue », à l'instar de *petit chat est* (cf. Salem 1987 : 51). On notera par ailleurs quelques tentatives récentes de dégager des schémas syntaxiques de groupes de segments répétés et de s'interroger sur leur fonctionnement discursif ; on mentionnera les segments possessifs (Cislaru *et al.* 2013), ou encore les constructions autour du verbe *pouvoir* (Née *et al.* 2016). Ces démarches sont proches des travaux de Renouf et Sinclair (1991) sur les cadres collocationnels⁶, bien que leurs hypothèses fonctionnelles, peu présentes dans le courant de la grammaire des usages britannique, rappellent davantage les travaux de Longrée et Mellet (2013) sur les motifs, ou ceux de Biber *et al.* (2004) sur les paquets lexicaux ou encore, à l'oral, les bribes ou débuts de syntagmes (Blanche-Benveniste 2003).

- 23 Les jets textuels, identifiés comme des unités relevant de l'autosegmentation, n'ont pas encore fait l'objet de descriptions linguistiques systématiques, mais leur description lexico-sémantique et syntaxique est un passage incontournable afin de saisir leur statut linguistique. Il est moins évident de formuler des hypothèses fonctionnelles, compte tenu du point de vue en production, qui amplifie la coupure d'avec le destinataire propre à l'écrit en dehors des médiations connectées. Nous proposerons cependant plus bas quelques hypothèses concernant leur rôle dans le processus de textualisation.
- 24 En termes d'approche lexico-sémantique, il apparaît difficile de s'inspirer de la description des segments répétés, qui implique directement la dimension statistique de la fréquence, quasi-absente dans le cas des jets textuels, très peu récurrents. C'est davantage une analyse des champs lexicaux qui s'imposerait, en suivant l'hypothèse de thèmes et domaines facilitant la sémiotisation spontanée à l'écrit. Ainsi, comme nous le notons plus haut, la désignation de l'enfant et des membres de la famille se produit-elle régulièrement sous forme de jets textuels récurrents, tout comme la désignation de certains « diagnostics » (*son manque de travail*) ; ces cas de figure relèvent davantage de l'apprentissage ou de la reproduction de séquences idiomatiques. En revanche, il serait intéressant d'analyser le dictionnaire des fréquences du sous-corpus de jets textuels et de le comparer au dictionnaire du corpus de rapports finalisés., de regrouper les jets textuels nominaux autour de champs dénotationnels lorsque cela est possible (désignation de l'enfant, désignation des troubles), etc.
- 25 Notre corpus de rapports éducatifs fournit 5 709 jets textuels que nous avons analysés et comparés aux segments répétés issus du même corpus, et donc produits par les mêmes scripteurs. Nous avons constaté que 5 % seulement des jets textuels et des segments répétés partageaient la même forme et le même contenu⁷, ce qui nous conduit à questionner le statut des premiers.

Tableau 1. Caractéristiques générales du corpus de rapports éducatifs enregistrés en temps réel

	Rapports										
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Total
Nbre de sessions	13	5	13	8	9	8	4	9	9	2	80
Nbre de mots dans les rapports finaux	3223	7400	3144	3356	2841	2878	892	1239	2776	1372	29121
Nbre de segments répétés	796	628	1432	934	983	1060	113	110	754	436	7246
Nbre de jets textuels	1092	441	373	949	868	563	493	394	474	62	5709

- 26 D'un point de vue syntaxique, on constate que moins de la moitié des jets (43 %) emprunte un format syntaxique valide, *i. e.* qui corresponde aux constituants de phrase. Faut-il pour autant négliger les 57 % restants⁸, qui représentent un pourcentage important de productions langagières ? Nous citerons ici opportunément Delofeu (2013) :

[...] je voudrais montrer que la notion de « fragment », terme nouvellement introduit dans les études linguistiques, n'est pas une notion descriptive, mais une

façon allusive de constater l'écart qu'il y a entre la représentation que certains linguistes se font de la langue et la réalité des productions langagières. (Deulofeu 2013 : 157)

- 27 Se plaçant sur le terrain de l'oral, Deulofeu (2013 : 166) postule l'irrégularité de l'interface syntaxe-lexique et plus largement sémantique : « [...] nous faisons l'hypothèse que la réalisation d'une place syntaxique quelconque peut se faire soit par un lexème soit par une absence de lexique, qu'elle soit codifiée en langue ou le simple effet d'un 'on line mechanism' », et note que la rupture de cohésion syntaxique ne remet point en cause le respect de la cohérence sémantique. La segmentation spontanée en jets textuels se rapproche ainsi à certains égards des bribes décrites par Blanche-Benveniste (1987) ou encore de la grammaire de l'oral envisagée par Berrendonner (2004), ou encore Brazil (1995) : une grammaire linéaire à unités non hiérarchisées et à dynamique incrémentale⁹.

2. Le segment comme unité de fonctionnement

- 28 La description lexicale, sémantique ou syntaxique des segments devrait pouvoir rendre compte du rôle que lesdits segments assument au niveau du texte, des apports en termes de lisibilité, interprétabilité, ou encore de performativité. Il nous semble que c'est autour des concepts de fréquence (récurrence, répétition, etc.) et de pré-construction (préfabriqué, sens partagé, etc.) que se cristallisent les questionnements fonctionnels.

2.1 Fréquence, répétition, routines : quel impact fonctionnel ?

- 29 Au niveau des discours, les observations de Lafon (1984 : 75) concernant l'absence de corrélation entre régularité et répétition permettent déjà d'entrevoir les difficultés que soulève la fréquence dans le cadre des analyses linguistiques des textes :

Ainsi, la répétition massive d'une forme, qu'elle soit fonctionnelle ou lexicale, se conjugue rarement avec une distribution régulière. Nous avons pu vérifier ce fait sur de nombreux autres textes. Plus une forme est souvent attestée dans un corpus, plus il est périlleux de penser que la connaissance de sa seule fréquence permet de bâtir un modèle pour prévoir ses apparitions dans un texte analogue plus court ou plus long. La valeur élevée d'une fréquence n'est pas un gage de sa stabilité. C'est même, le plus souvent, l'inverse qui est vrai.

- 30 La répétition, à laquelle on attribue un sens ou, du moins, des effets de sens au niveau du discours, rend visibles et pertinentes les unités de type segments répétés. Le maillage des segments se constitue ainsi en trame sémantique ou textuelle à effet de réticularité (cf. Legallois 2006) que l'outil aide à saisir et à interpréter. La répétition peut ainsi s'interpréter comme relevant d'un effet sémantique localisé et non prédictible *a priori*, à la différence de la régularité, qui serait sous-tendue par des principes plus généraux, s'appliquant de manière attendue à grande échelle. Comme le souligne Hyland (2011 : 108), la typicalité des choix linguistiques mis au jour par l'analyse du discours montre l'habituel et le déviant, le globalement sous-jacent et les préférences locales. Plus spécifiquement, il s'agit de prendre en compte l'opposition entre pratiques collectives, où s'ancrent davantage les régularités, et usages contextuels/individuels, qui font émerger les répétitions. Il nous semble que la

corrélation entre fréquence et conditions de production rend bien compte de la charge sémantique que porte la fréquence :

En première approximation les fréquences peuvent être considérées comme une manifestation des conditions de production [conditions de production doit s'entendre ici dans un sens élargi qui comprend des facteurs intrinsèques au discours] du discours sans être pour cela déterminées par elles. (Lafon 1984 : 50)

- 31 Les faiblesses de la notion de fréquence et de son exploitation statistique ont été rappelées récemment par plusieurs auteurs (cf. notamment Blumenthal-Dramé 2013 : 33-39 ; Stefanowitsch et Gries 2003). Bley-Vroman (2002) montre par exemple comment une approche sémantique du langage, prenant appui sur le principe d'association des formes et des sens, permet d'expliquer la production et l'acquisition de structures langagières tout en tenant compte des effets de la fréquence sur les savoirs linguistiques. Ainsi, les segments ayant un statut idiomatique préalable au corpus d'analyse constituent souvent des occurrences uniques au niveau d'un corpus cerné : si les locuteurs savent généralement reconnaître le statut idiomatique d'une séquence comme *donner du fil à retordre*, ce segment a peu de chances d'être récurrent dans de nombreux corpus d'analyse, sauf si on fait appel à des bases de données. En revanche, au niveau des usages linguistiques, les segments idiomatiques sont préférentiellement reproduits à l'identique ; la répétition est ainsi compilée sur de larges empan temporels et enregistrée dans des thésaurus écrits ou oraux plutôt qu'identifiée en corpus. C'est notamment en tant que catalyseur de l'unité sens-forme, du figement au sein de la communauté linguistique, qu'intervient la répétition. L'unité sens-forme apparaît comme un produit collectif, validé par des pratiques langagières réitérées, tandis que les usages linguistiques spécifiques et la reproduction des segments à l'identique relèvent du domaine de l'individuel, qui s'inscrit de cette manière dans une communauté dont il partage les habitus. Se pose alors la question du rapport entre répétition et apprentissage à l'échelle des usages individuels, où l'automatisme de reproduction à l'identique serait davantage le résultat d'un apprentissage ciblé, sélectionné et régulé que le produit d'actualisations répétées individuellement¹⁰. Les notions de régularité et de stabilité sont donc à convoquer en parallèle pour toute approche de la fréquence.
- 32 Corrélé à la notion de fréquence qui en est un des mécanismes sous-tendant, le principe de routinisation se caractérise par la prédictibilité et est omniprésent dans plusieurs approches du corpus s'inscrivant notamment dans le cadre des linguistiques de l'usage (*usage-based, bottom-up*, etc. Schmid 2014, Sitri et Tutin 2016). Or, la répétition et la routinisation ont en commun l'actualisation d'une charge sémantique. Du point de vue de la théorie de l'enracinement et de la conventionalisation (*entrenchment and conventionalization*, Schmid 2014¹¹) la routinisation configure les associations entre structures linguistiques, sens, schémas, etc. durables et les rend fonctionnelles en langue et/ou en discours, en sous-tendant ainsi les régularités.
- 33 Dans cette optique, certains jets textuels pourraient être considérés comme des produits d'un processus de routinisation en usage. Les jets textuels étant peu récurrents dans notre corpus – peu de jets textuels y apparaissant deux fois ou plus –, ce n'est pas l'identité formelle de séries de séquences qui permettrait de poser la question de la routinisation ou de la régularité (à la différence des segments répétés). C'est la spontanéité d'actualisation, corrélée à quelques régularités syntaxiques dont il sera question plus bas, qui interroge le rapport des jets textuels à la régularité et à la stabilité. Qu'est-ce qui permet la production spontanée des jets textuels, quelles

routines scripturales, quelles régularités génériques ? Il s'agit alors d'identifier les paramètres sémantiques qui seraient attachés à une routinisation préalable à la production des jets. La répétition ou la fréquence étant non pertinentes pour la définition des jets textuels produits spontanément à l'écrit, faudrait-il alors complètement dissocier fréquence et pré-construit ? Si tel est le choix qui s'impose, comment justifier la pré-construction en dehors de toute répétition ou routinisation ?

- 34 C'est dans l'optique de ce questionnement que sont discutés ci-après les concepts de pré-construction, pré-fabrication, lexicalisation, et les implicites sémantiques qu'ils charrient.

2.2 Pré-construction, lexicalisation, sens

2.2.1 Le pré-construit à l'aune du processus et du produit d'écriture

- 35 Comme le soulignent Gülich et Krafft (1997 : 256), les structures préformées peuvent remplir une fonction sociale, en renvoyant à un savoir partagé et en renforçant ainsi la cohésion d'une communauté. Wray et Perkins (2000 : 11) rappellent également que l'intérêt des préfabriqués est celui du moindre effort, dû à l'emploi de ce qui est attendu (*the expected*). Les auteurs cités se situent visiblement du point de vue de la réception. Or, au niveau du processus de rédaction, c'est plutôt le contraire qui se passe : l'effort – avec les réécritures, ajustements, etc. – vise à atteindre ce qui est attendu une fois le texte finalisé. Le faible degré de similitude entre jets textuels et segments répétés (5 %) montre par exemple que ces derniers sont rarement pré-construits, au sens où ils sont rarement produits dans leur forme finalisée. En diminuant le coût des traitements de production, les préfabriqués donneraient lieu à des jets textuels.
- 36 Nous avons ainsi entrepris d'identifier le niveau de préconstruction des jets textuels. Dans une première tentative de classification des jets textuels, nous avons constaté que les repères lexico-idiomatiques ou syntaxiques habituels étaient peu pertinents, du fait de la faible récurrence et de la non-saturation syntaxique prégnante des jets (voir plus haut). Ainsi, les jets textuels que nous avons recueillis révèlent relativement peu de séquences candidates au statut de préfabriqué, tels que les connecteurs produits sous forme de jets récurrents (*De plus, En effet, à chaque fois, à ce sujet, au quotidien, en permanence*) et les clichés génériques (*Nous observons, en milieu ouvert, droits de visite et d'hébergement*). Nous nous sommes alors tournés vers une analyse constructionnelle, en cherchant les récurrences morphosyntaxiques et les éventuelles formes-pôles fréquentes, qui seraient révélatrices de pré-construits sous forme de patrons. Nous avons notamment identifié deux types de jets textuels qui semblent correspondre *grasso modo* à des cadres collocationnels productifs au niveau du processus de rédaction : les jets textuels coordinatifs (ex. 1, 2) et les jets textuels s'organisant autour d'un connecteur (ex. 3). Ces cadres collocationnels allient des contraintes combinatoires déterminées par la nature des noyaux (conjonctions de coordination, connecteurs) à des contours lâches, qui restent ouverts à des enchaînements syntagmatiques au sein du même jet textuel. Nous allons nous appuyer sur ces catégories de jets textuels pour discuter des niveaux de pré-construction d'une part, du rôle des segments d'autre part.

2.2.2 Les jets textuels coordinatifs comme cadres collocationnels pré-construits

- 37 La conjonction *et* est attestée en moyenne¹² dans presque 7 % des jets textuels produits lors du processus d'écriture des dix rapports éducatifs. Un seul dossier, qui contient 62 jets textuels seulement, n'atteste pas de jet en *et* (cf. tableaux 1 et 2) :

- (1) a) *et les résultats*
 b) *ces états d'agitation régulières et*
 c) *la vie quotidienne et ordinaire au sein d'une famille.*

Tableau 2. Nombre et pourcentage de jets contenant la conjonction *et*

	Rapports										Total
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	
Jets	68	50	20	79	57	50	27	18	24	0	393
en ET	6.2 %	11 %	5.4 %	8.3 %	6.6 %	8.9 %	5.5 %	4.6 %	5.1 %	-	

- 38 Nous avons été en mesure de proposer une classification en termes de constituants relationnels pour 183 occurrences du patron *X et Y* (sur un total de 225 occurrences¹³). Un tiers des jets met en relation des noms ou groupes nominaux :

Tableau 3. Constituants du patron *X et Y*

	Types de constituants					Total
	GN+et+GN N+et+N	GP+et+GP	GV+et+GV V+et+V	Adj+et+Adj ou équivalent	Propositions	
Exemple	<i>entre père et fille</i>	<i>l'endroit de l'équipe et des autres jeunes</i>	<i>a 12 ans et subit beaucoup de transformations corporelles</i>	<i>la vie quotidienne et ordinaire au sein d'une famille</i>	<i>pour tenter d'élaborer ce qui lui pose problème et ce qui l'envahit</i>	
Nombre	64	34	16	20	35	183 / 225
Pourcentage	34.7 %	15.1 %	7.1 %	8.9 %	15.6 %	81.3 %

- 39 La préférence des coordinations nominales peut-elle constituer un argument en faveur d'une pré-construction ? Au moins deux hypothèses s'offrent à nous. L'une s'attache au système et prédit que le coordinatif *et* relie souvent des noms, comme nous avons pu le voir en sondant d'autres corpus (presse, romans policiers) avec *Le Trameur*¹⁴. Au vu des régularités structurelles observées, un certain degré de pré-construction faciliterait l'actualisation de jets coordinatifs de l'un ou l'autre type. L'autre s'attache au genre et s'appuie sur le constat que les discours spécialisés marquent une préférence pour la nominalisation (cf. Kocourek 1991). Bien que s'appuyant sur ce qui est devenu une sorte de postulat définissant les discours spécialisés, cette deuxième hypothèse reste invérifiable empiriquement dans les limites du présent travail, compte tenu de l'absence de données différentielles concernant les spécificités des discours spécialisés en français. L'observation qualitative des jets *N+et+N* dans notre corpus montre que, malgré quelques nominalisations à proprement parler (*sa prise d'indépendance et ses relations*), ceux-ci renvoient plutôt au champ référentiel de la famille et de la vie

quotidienne (*son linge propre et son linge sale*). Au niveau générique, la pré-construction de ces jets pourrait ainsi être sous-tendue par la dénotation.

- 40 En observant les jets coordinatifs en *mais* (58 occurrences, dont 36 binaires *X mais Y*),
 (2) a) *été chagriné mais pas abattu*
 b) *mais qu'il en était victime*
- 41 on voit que le sens de la conjonction peut attirer certains constituants et en éliminer d'autres, les noms et les verbes n'étant point représentés dans ces jets. D'une manière générale, les jets binaires autour de *mais* sont peu homogènes syntaxiquement et restent réticents aux critères de classification que nous nous sommes donnés dans le cadre du présent travail (seuls 36,1 % des jets ont pu être catégorisés) :

Tableau 4. Nombre et pourcentage de jets textuels présentant un patron *X mais Y*

	Types de structures					Total
	GN+et+GN N+et+N	GP+et+GP	GV+et+GV V+et+V	Adj+et+Adj ou équivalent	Propositions	
Exemple		<i>régulièrement au travers des moments de retour en week- ends ,mais aussi lors de rendez-vous</i>		<i>[...] se montre plutôt indifférent mais pas attaquant. [...]*</i>	<i>lle semble avoir beau- coup d'amis mais les rela- tions sont très fluctuantes</i>	
Nombre	-	1	-	2	10	13
Pourcentage	-	2.8 %	-	5.6 %	27.8 %	36.1 %

* Pour les deux occurrences des jets *Adj+mais+Adj* on notera la négation du second adjectif, négation qui marque la réponse à une attente interprétative confirmant ou infirmant certains clichés psy.

- 42 Les disparités et les préférences structurelles des jets coordinatifs, corrélées à la proportion de ces derniers suggèrent que c'est plus particulièrement un type de relation référentielle et discursive qui est pré-construit.

2.2.3 Jets coordinatifs et segments dans les versions finalisées des rapports

- 43 Nous avons observé les textes finalisés des rapports éducatifs, pour tenter une articulation entre les segments du processus et leur pendant stabilisé au niveau du produit. On constate que la conjonction *et* est relativement fréquente (596 occurrences, soit 2 % des formes), et que ses constructions se répartissent de manière équilibrée sur l'ensemble de la surface textuelle, sans préférence marquée pour les différentes rubriques (histoire familiale, scolarisation, santé, relations avec les parents, conclusions...) ou pour une topologie spécifique. L'étude des cooccurrences de la conjonction *et* dans les rapports finalisés semble indiquer un haut degré de prospection. Ainsi, *entre* et *Direction*¹⁵ sont les formes qui enregistrent la spécificité la plus élevée parmi les cooccurrents de *et* (co-fréquence 28/33 et 19/20 respectivement). La préposition *entre* préfigurerait la relation coordinative (colligation, au sens de patron syntaxique), en s'associant d'ailleurs souvent à des termes relationnels ou des termes dénotant une dynamique impliquant deux points de repère (collocation) :

Tableau 5. Exemples de concordances prospectives des cooccurrents *entre* et *et*

Lexique relationnel	Lexique dénotant un changement ou une dynamique
<p>Les difficultés relationnelles <u>entre</u> A. et sa mère ont motivé le au domicile. Alors que les relations <u>entre</u> Madame O. <u>et</u> A. ont semblé s'améliorer</p> <p>Nous tentons de faire des liens <u>entre</u> les violences qu'il a subies <u>et</u> la mise en place de rencontres progressives <u>entre</u> D. et son père en tenant compte E. ne fait pas encore les liens <u>entre</u> les inquiétudes transmises par sa mère <u>et</u> 'arrivée des petits peut modifier les relations <u>entre</u> F. <u>et</u> son père. En effet, autre chose, refusant de faire des liens <u>entre</u> la violence vue, vécue <u>et</u> la</p>	<p>recherche en permanence le conflit. Elle oscille <u>entre</u> provocation <u>et</u> excitation, ce qui rend lien avec ses conduites addictives. Madame oscille <u>entre</u> le déni <u>et</u> la reconnaissance de ses</p> <p>La période de transition <u>entre</u> le service de pédopsychiatrie, très contenant par le dispositif proposé, <u>et</u> pu montrer des inquiétudes face à l'organisation notamment lors des passage <u>entre</u> les différents lieux d'accueil (famille d'accueil <u>et</u> ITEP).</p>

- 44 Nous pouvons nous référer ici aux notions de prédiction ou de prospection, définies par Sinclair (2004), et considérer les jets textuels coordinatifs comme des déclencheurs de schémas pré-construits aussi bien au niveau du processus de rédaction que dans le texte finalisé. En amont, d'autres déclencheurs comme les constructions en *entre* créent des attentes relationnelles et préparent ainsi la place aux jets textuels.
- 45 Dans une optique différente, mais tout aussi instructive, on notera que les cooccurrents du coordonnant *mais* dans les textes finalisés comptent (voir l'annexe pour les détails) :
- une proportion élevée de verbes modaux (sur 11 verbes on en compte 5 occurrences : *peut, pouvait, semble, souhaitait, devrait*), attributifs (*reste, semble*) et des verbes de parole ou intellectuels (*verbaliser, écrire, comprendre*) ;
 - des marqueurs de renforcement ou d'insuffisance, dont certains pourraient fonctionner comme des marqueurs concessifs (*aussi, sans, moins, pas*) ;
 - des noms marquant une rupture (événements, agressivité).
- 46 Dès le processus de production écrite, on compte déjà, parmi les jets textuels en *mais*, 13 occurrences à valeur de renforcement ou d'atténuation contenant *mais aussi, mais également, mais à la fois, mais en même temps*, dont 3 occurrences de *mais aussi*. S'ajoutent cinq occurrences d'emploi polémique (ex. : *se montre plutôt indifférent mais pas attaquant*). On peut ainsi confirmer la présence de ces schémas à potentiel argumentatif dès les premiers jets des rapports.

2.2.4 Les jets textuels autour d'un connecteur et les routines scripturales

- 47 Enfin, nous interprétons les jets textuels construits autour de connecteurs (3) à la lumière des propositions de Perrin (2011 : 83), selon lequel le figement, qui concerne « l'établissement graduel de tout ce qui est pré-construit dans le langage », serait à appréhender d'abord à « un niveau sémantico-pragmatique, indépendamment des restrictions syntaxiques qu'il fait subir ». L'auteur considère que l'ensemble des marques énonciatives, discursives et même grammaticales relèvent d'une forme de figement « insensible au sujet parlant », ce sont des marques spécialisées conventionnelles (Perrin 2011 : 88-89). Ainsi, en assimilant les connecteurs, à côté des

formules énonciatives, des préfixes modaux ou performatifs, ou encore des adverbes de phrase, à des formules, il précise que les « formules sont des symptômes conventionnels et donc des formes linguistiques, issues diachroniquement d'un comportement stéréotypé associé à une forme de routine énonciative et finalement de figement formulaire » (Perrin 2011 : 87).

(3) a) *proche du domicile familiale. Bienque séparer*

b) *s'interrogeait sur la judiciarisation. Ainsi, il*

c) *été reçu également, évoquant ses difficultés avec beaucoup de lucidité et notamment le bienfait des soins mis en place. Parallèlement, elle pouvait évoquer le vide d'être à distance des organisations familiales.*

d) *Dernièrement, nous avons retrouvé des armes factices de p*

e) *En attendant lors des entretiens il mélange le linge propre et le linge sale ce qui en terme*

f) *Ainsi, il ramène des tableaux ou*

- 48 Nous retenons notamment la notion de routine interprétative déclenchée par l'énonciation de l'expression (Perrin 2011 : 88), à laquelle nous ajouterions un pendant processuel : l'inscription d'un connecteur est sous-tendue par une routine scripturale et engendre des formats-types aux niveaux sémantique, syntaxique et textuel.

Conclusions

- 49 La définition et la caractérisation des jets textuels, segments écrits produits spontanément entre deux pauses, ne peut faire l'économie d'une discussion des critères définitoires s'appliquant traditionnellement à la définition des figements, segments textuels récurrents et autres schémas collocationnels. Nous avons pu constater que des critères définitoires habituellement convoqués comme la fréquence, la récurrence, la régularité, la stabilité, la pré-construction, l'automatisme, la mémorisation s'appliquent de manière sélective aux différents types de segments : ainsi, les segments répétés sont caractérisés par leur fréquence mais pas nécessairement par l'automatisme ou la stabilité ; les figements ne sont pas nécessairement récurrents en synchronie ; les jets textuels ne répondent pas aux critères de fréquence, etc.
- 50 Ainsi, en nous penchant sur le cas spécifique des jets textuels, nous avons remarqué que, loin de se résumer à des traits pertinents/non pertinents, chacun de ces critères peut intervenir, bien que de manière non linéaire, dans leur définition. La fréquence et la récurrence sont caractéristiques de schémas syntaxiques à contours lâches et à noyaux-pôle à fonction relationnelle et prospective au niveau du texte. Des colligations comme la relation coordinative (*et, mais*) et des collocations contraignant les types de termes relationnels (noms, verbes, adjectifs, etc. avec, à l'intérieur de ces catégories, des champs sémantiques préférentiels) actualisés dans les jets textuels suggèrent que les régularités et pré-construits s'appliquent à deux niveaux différents, la syntaxe et les domaines dénotationnels, en dépassant ainsi le principe de récurrence lexico-formelle propre notamment aux segments répétés. S'il est difficile de se prononcer sur le caractère préfabriqué des jets textuels – comment évaluer leur degré de stabilité, par exemple, en l'absence de toute répétition ? – il apparaît en revanche que i) les cadres collocationnels coordinatifs sont pré-construits ; ii) ils contraignent dans une série de cas le choix des parties de discours à actualiser ; iii) des schémas d'organisation textuelle sont préfigurés dans/par les jets textuels coordinatifs. On distinguera ainsi deux niveaux de pré-constructions : les jets textuels coordinatifs et ceux organisés autour de connecteurs seraient des pré-construits en vertu de leurs schémas réguliers,

mais, de par leur fonctionnement prospectif, ils représentent des pré-construits du schéma argumental du texte. On émet l'hypothèse d'un schéma pré-construit sous-jacent attaché aux formules, celles-ci assumant des rôles d'amorces ou de pôles de jonction de contenus ou de constructions et facilitant ainsi la production sous forme de jets textuels. La description sémantique et thématique des jets textuels s'organisant autour d'un connecteur notamment et l'analyse de leur statut rédactionnel (jets de production *versus* jets de révision) nous permettront de vérifier cette hypothèse dans des travaux ultérieurs.

- 51 La spontanéité de production des jets textuels peut indiquer, jusqu'à preuve du contraire, un certain degré d'automatisme, qui se traduirait par une récupération en mémoire à long terme plutôt que par la mise en œuvre de processus de génération de texte. Cependant, la question de la mémorisation reste entière, dans la mesure où peu de figements lexicaux sont repérés parmi les jets textuels étudiés. Les cadres collocationnels constituent bien des séquences mémorisées, mais cela ne pourrait suffire à assurer la spontanéité de production des jets, bien qu'ils puissent la faciliter. On peut ainsi évoquer le rôle de la mémoire de travail ou à court terme dans la récupération de patrons figés, le choix du lexique étant en quelque sorte préconisé par le patron syntaxique sélectionné.
- 52 Les jets textuels s'inscrivent ainsi dans un tissage de segments de natures diverses, qui semblent représenter, sous des angles différents, des moules langagiers, du toujours-déjà-là qui sous-tendent les pratiques langagières en tous genres.

BIBLIOGRAPHIE

- Alves R.A., Branco M., Castro S. L. et Olive T. (2012). « Effects of handwriting skill, output modes and gender of fourth graders on pauses, written language bursts, fluency and quality », in V. W. Berninger (éd.), *Past, present, and future contributions of cognitive writing research to Cognitive Psychology*. New York : Psychology Press.
- Alves R. A., Castro S. L. et Olive T. (2012). « Transcription skill constrains bursts of language production », in M. Torrance, D. Alamargot, M. Castello, F. Ganier, O. Kruze, A. Manger, L. Tolchinsky et L. Van Waes (éd.), *Learning to write effectively : current trends in European research*. UK : Esmerald Group Publishing, p. 27-29.
- Alves R. A., Castro S. L., Sousa L. et Strömqvist S. (2007). « Typing skill and pause-execution cycles in written composition », in M. Torrance, L. Van Waes, D. Galbraith (éd.) *Writing and cognition, research and applications*. Dordrecht : Elsevier Sciences Publishers, p. 55-65.
- Alves R. A., Limpo T., Fidalgo R., Carvalhais L., Pereira L. Á., et Castro S. L. (2016). « The Impact of Promoting Transcription on Early Text Production : Effects on Bursts and Pauses, Levels of Written Language, and Writing Performance ». *Journal of Educational Psychology* 108 (5) : 665-679.
- Anscombe J.-C. (2011). « Figement, idiomaticité et matrices lexicales », in J.-C. Anscombe et S. Mejri (éd.) *Le figement linguistique : la parole entravée*. Paris : Honoré Champion, p. 17-40.

- Arrivée M. (1998). « Unité linguistique et unité sémiologique chez Ferdinand de Saussure », in G. Quiroz, I. Berthoud-Papandropoulou, E. Thommen, C. Vogel (éd.) *Les unités discursives dans l'analyse sémiotique : la segmentation du discours*. Berne et al. : Peter Lang.
- Baaijen Veerle M., Galbraith D. et de Glopper K. (2012). « Keystroke Analysis : Reflections on Procedures and Measures ». *Written Communication* 29 (3) : 246-277.
- Berrendonner A. (2004). « Grammaire de l'écrit vs grammaire de l'oral : le jeu des composantes micro- et macro-syntaxiques », in A. Rabatel (éd.) *Interactions orales en contexte didactique. Mieux (se) comprendre pour mieux (se) parler et pour mieux (s')apprendre*. Lyon : IUFM et Presses universitaires de Lyon, p. 249-264.
- Biber D., Conrad S., Cortes V. (2004). « If you look at... : Lexical bundles in university teaching and textbooks », *Applied Linguistics* 25 (3) : 371-405.
- Blanche-Benveniste C. (1987). « Syntaxe, choix du lexique et lieux de bafouillage », *DRLAV* 36-37 : 123-157.
- Blanche-Benveniste C. (2003). « La naissance des syntagmes dans les hésitations et les répétitions du parler » in J.-L. Araoui (éd.), *Le sens et la mesure. Hommages à Benoît de Cornulier*. Editions Honoré Champion, p. 40-55.
- Bley-Vroman R. (2002). « Frequency in production, comprehension, and acquisition », *Studies in Second Language Acquisition* 24 : 209-213.
- Brazil D. (1995). *A Grammar of Speech*. Oxford : Oxford University Press.
- Bybee J. (2010). *Language, Usage and Cognition*. Cambridge : Cambridge University Press.
- Chenoweth A. N. et Hayes J. R. (2001). « Fluency in writing », *Written Communication* 18 : 80-98.
- Chenu F., Pellegrino F., Jisa H. et Fayol M. (2014). « Interword and intraword pause threshold in the writing of texts by children and adolescents : a methodological approach ». *Frontiers in Psychology* 5 : 182.
- Cislaru G., Pugnière-Saavedra F., Sitri F. (éd.) (2008). « Analyse de discours et demande sociale », *Les Carnets du Cediscor* 10.
- Cislaru G., Sitri F. et Pugnière-Saavedra F. (2013). « Figement et configuration textuelle : les segments de discours répétés dans les rapports éducatifs », in C. Bolly et L. Degand (éd.), *Across the Line of Speech and Writing Variation. Corpora and Language in Use*, 165183. Louvain-la-Neuve : Presses universitaires de Louvain.
- Connelly V., Dockrell J., Walter K. et 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.
- Deulofeu J. (2013). « De l'inutilité de la notion de "fragment" pour la description des énoncés "fragmentés" », in P. Hadermann, M. Pierrard, A. Roig, D. V. Raemdonck (éd.) *Ellipse & fragment. Morceaux choisis*. Bruxelles et al. : Peter Lang, p. 157-179.
- Edmonds A. (2013). « Une approche psycholinguistique des phénomènes phraséologiques : le cas des expressions conventionnelles », *Langages* 189 : 121-138.
- Fenoglio I. (2007). « Du texte avant le texte. Formes génétiques et marques énonciatives de prévisions textualisantes », *Langue française* 155 : 8-34.
- Goldberg A. (2006). *Constructions at Work : the nature of generalization in language*. Oxford : Oxford University Press.

- Grunig B.-N. (1997). « La locution comme défi aux théories linguistiques : une solution d'ordre mémoriel ? », in M. Martins-Baltar (éd.) *La locution entre langue et usages*. Paris - Fontenay-Saint-Cloud : ENS Éditions, p. 225-240.
- Gülich E., Krafft U. (1997). « Le rôle du préfabriqué dans les processus de la production discursive », in M. Martins-Baltar (éd.) *La locution entre langue et usages*. Paris - Fontenay-Saint-Cloud : ENS Éditions, p. 241-276.
- Hunston S. (2014). « Pattern grammar in context », in T. Herbst, H.-J. Schmid, S. Faulhaber (éd.) *Constructions, Collocations, Patterns*. Berlin - Boston : De Gruyter Mouton, p. 99-119.
- Hyland K. (2011). « Looking through corpora into writing practices », in V. Viana, S. Zyngier, G. Barnbrook (éd.) *Perspectives on Corpus Linguistics*. Amsterdam - Philadelphia : John Benjamins, p. 99-114.
- Kocourek R. (1991). *La langue française de la technique et de la science, vers une linguistique de la langue savante*. Wiesbaden : Oscar Brandstetter Verlag.
- Lafon P. (1984). *Dépouillements et statistiques en lexicométrie*. Genève : Slatkine - Champion.
- Lafon P. et Salem A. (1983). « L'inventaire des segments répétés d'un texte », *Mots* 6 : 161-177.
- Lebart L., Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Legallois D. (2006). « Des phrases entre elles à l'unité réticulaire du texte », *Langages* 163 : 56-70.
- Legallois D. (2012). « La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? », *Corpus* 11 [En ligne], mis en ligne le 21 juin 2013, consulté le 22 juin 2016. <http://corpus.revues.org/2202>.
- Legallois D., Tutin A. (éd.). (2013). « Vers une extension du domaine de la phraséologie », *Langages* 189.
- Leijten M. et Van Waes L. (2013). « Keystroke Logging in Writing Research : Using Inputlog to Analyze and Visualize Writing Processes », *Written Communication* 30 : 358-392.
- Loiseau S. (2015). « Les différentes formes de la fréquence textuelle : proposition d'inventaire », *Langages* 197 : 1-17.
- Longrée D., Mellet S. (2013). « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours », *Langages* 189 : 65-79.
- Mahrer R., Nicollier Sarailon V. (2014). « Les brouillons font-ils texte ? Le cas des plans pré-rédactionnels de C. F. Ramuz », in J.-M. Adam (éd.), *Faire Texte. Frontières textuelles et opérations de textualisation*. Besançon : Annales littéraires de l'Université de Franche-Comté, p. 223-305.
- Mayaffre D. (2007). « L'analyse de données textuelles aujourd'hui : du corpus comme une urne au corpus comme un plan. Retour sur les travaux actuels de topographie/topologie textuelle (partie I) », *Lexicometrica* 9. <http://-lexicometrica.univ-paris3.fr/numspeciaux/special9/mayaffre.pdf>.
- Née E., Sitri F., Veniard M. (2016). « Les routines, une catégorie pour l'analyse de discours : le cas des rapports éducatifs », *Lidil* 53 : 71-93.
- Olive T. (2011). « Working memory in writing », in V. W. Berninger (éd.), *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*. New York : Psychology Press, p. 485-503.
- Olive T. (2014). « Toward an Incremental and Cascading Model of Writing : A review of research on writing processes coordination », *Journal of Writing Research* 6 : 173-194.

- Olive T., Cislaru G. (2015). « Linguistic forms at the -process- product interface : analyzing the linguistic content of bursts of production », in G. Cislaru (éd.) *Writing(s) at the crossroads : the process/product interface*. Amsterdam : John Benjamins, 99-123.
- Perrin L. (2011). « Figement, énonciation et lexicalisation citative », in J.-C. Anscombre et S. Mejri (éd.) *Le figement linguistique : la parole entravée*. Paris : Honoré Champion, p. 81-94.
- Polguère A. (2007). « Soleil insoutenable et chaleur de plomb : le statut linguistique des greffes collocationnelles », in M.-C. L'Homme et S. Vandaele (dir.) *Lexicographie et terminologie : compatibilité des modèles et des méthodes*. Ottawa : Les Presses de l'Université d'Ottawa, p. 247-291.
- Renouf A., Sinclair J. (1991). « Collocational frameworks in English », in K. Aijmer et B. Altenberg (éd.) *English Corpus Linguistics : Studies in Honour of Jan Svartvik*. London : Longman, p. 128-143.
- Salem A. (1987). *Pratique des segments répétés. Essai de statistique textuelle*. Paris : INALF - Klincksieck.
- Schmale G. (2013). « Qu'est-ce qui est préfabriqué dans la langue ? – Réflexions au sujet d'une définition élargie de la préformation langagière », *Langages* 189 : 27-45.
- Schmid H.-J. (2014). « Lexico-grammatical patterns, pragmatic associations and discourse frequency », in T. Herbst, H.-J. Schmid, S. Faulhaber (éd.) *Constructions, Collocations, Patterns*. Berlin - Boston : De Gruyter Mouton, p. 239-293.
- Sinclair J. (2004). *Trust the Text. Language, Corpus and Discourse*. London : Routledge.
- Sinclair J., Mauranen A. (2006). *Linear Unit Grammar : Integrating Speech and Writing*. Amsterdam - Philadelphia : John Benjamins.
- Sitri F., Tutin A. (éd.). (2016). *Phraséologie et genres de discours, Lidil* 53.
- Stefanowitsch A., Gries S. (2003). « Collostructions : investigating the interaction of words and constructions », *International Journal of Corpus Linguistics* 8(2) : 209-243.
- Tamba I. (2011). « Sens figé : idiomes et proverbes », in J.-C. Anscombre et S. Mejri (éd.) *Le figement linguistique : la parole entravée*. Paris : Honoré Champion, p. 109-126.
- Wengelin Å. (2014). « Temps et pause dans l'écriture au clavier », in C. Leblay et G. Caporossi (éd.), *Temps de l'écriture. Enregistrements et représentations*, 97124. Paris : L'Harmattan.

ANNEXES

Cooccurents de *mais* dans les versions finalisées des rapports éducatifs.

```

Pôle : mais
Fq : 68
Co-Freq : 2 | Seuil : 3
ARC Couleur => bleu: (3<=Sp<=5) | vert: (5<Sp<=10) | orange: (10<Sp<=50) | rouge: (50<Sp)
ARC Epaisseur -> 1: (1<=NbContexte<=20) | 2: (20<NbContexte<=40) | 3: (40<NbContexte<=60) | 4: (60<NbContexte)
ARC label : cofreq(Sp)(nbContexte)

```

Pôle	Cooccurrent▲	Fq (Cooccurrent)	co-freq	specif	contextes
mais	accompagner	5	3	4	3
mais	agressivité	3	2	3	2
mais	aussi	33	9	5	9
mais	choix	10	3	3	3
mais	CHRS	5	2	3	1
mais	comprendre	11	3	3	3
mais	devrait	3	2	3	2
mais	écrire	3	2	3	1
mais	événements	3	2	3	2
mais	envisagé	4	2	3	2
mais	il	133	15	3	11
mais	Jimmy	5	2	3	2
mais	moins	26	5	3	5
mais	parfois	24	5	3	5
mais	pas	178	20	3	19
mais	passés	5	2	3	2
mais	PECHE	10	4	4	4
mais	père	72	9	3	8
mais	peut	89	12	3	11
mais	physique	5	2	3	2
mais	pour	279	25	3	23
mais	pouvait	14	4	3	3
mais	prendre	18	4	3	4
mais	qu'ils	5	2	3	2
mais	rappelé	4	2	3	2
mais	reste	30	5	3	5
mais	sans	29	5	3	4
mais	semble	35	6	3	6
mais	son	320	28	3	23
mais	souhaitait	3	2	3	2
mais	structure	4	2	3	2
mais	Toussaint	3	2	3	2
mais	verbaliser	5	2	3	2

NOTES

1. Legallois (2012) défend le même point de vue à propos des cadres collocationnels, par exemple : « Les cadres collocationnels sont des unités de langue, même s'ils ne sont pas habituellement mentionnés dans les descriptions linguistiques ».
2. Ainsi, les segments répétés (cf. Salem 1987) sont repérés avant tout en fonction de la récurrence des séquences graphiques dans un corpus ou dans une section du corpus.
3. « La ligne de démarcation entre la formule figée, locutionnelle, et la formule "normale", libre, n'est pas d'abord et avant tout un mode linguistique de construction grammaticale ou sémantique (même si, je l'ai souligné, certains schémas sont récurrents par analogie) mais – à certains égards plus aléatoire – *le fait qu'une inscription mémorielle s'est faite ou non. Toute forme linguistique est candidate. La mémoire prend ou ne prend pas, retient ou ne retient pas.* » (Grunig 1997 : 235).
4. Pour une présentation détaillée des rapports éducatifs et des enjeux discursifs qu'ils présentent, voir Cislaru *et al.* (2008).
5. Nous considérons, avec Fenoglio (2007) et Mahrer et Nicollier-Saraillon (2014) qu'il s'agit néanmoins de texte.
6. Les cadres collocationnels se définissent par des frontières constituées d'items grammaticaux prédéfinis (prépositions, conjonctions, déterminants, particules) et d'un cœur lexical sélectionné dans une liste de possibles. Ex. : *trop + Adj [beau, pauvre, intelligent] + pour*. Par extension, tout patron syntaxique comportant une partie prédéfinie et une partie « libre ».
7. Il s'agit notamment de connecteurs (*De plus, En effet, à chaque fois, à ce sujet, au quotidien, en permanence*), de paquets lexicaux ouvrant à droite vers une prédication (*Il est, X a, a été*), des désignations des enfants ou des membres de la famille.
8. Voir les jets textuels cités dans les séries d'exemples (1)-(3) pour une illustration de ces cas de figure.

9. Récemment, Sinclair et Mauranen (2006) ont tenté d'appliquer des propositions de D. Brazil aux textes écrits, en se situant cependant dans une perspective de réception et non de production, ce qui ne permet pas de rendre compte de la manière dont la syntaxe se déploie spontanément à l'écrit.
10. Notons cependant que l'apprentissage implique aussi la répétition.
11. « In sum, the major processes identified in the EC-model as being constitutive of a dynamic and adaptative model of language are the cognitive entrenchment-processes of **association**, **routinization** and **schema-formation** and the sociopragmatic conventionalization-processes of **communication** and **co-adaptation**. » (Schmid 2014 : 247).
12. Les écarts entre les dossiers peuvent se situer sur un empan assez large, cf. 4,6 % pour le dossier 8 versus 11 % pour le dossier 2. Ces spécificités, qui traduisent certainement des stratégies de rédaction différentes ne peuvent être traitées plus en détail dans les limites du présent article.
13. Les 42 occurrences restantes comptent des jets syntaxiquement non homogènes (*régulièrement et séparé* ; *au premier et 8,4/20 au second.*) ou des jets dont la composition lexicosyntaxique n'est pas aboutie (*de ce conflit et ce* ; *en cours et ains[i]*). La recherche de régularités potentielles dans cette série de jets reste à faire.
14. Outil d'analyse textométrique développé par Serge Fleury à l'Université Sorbonne nouvelle Paris 3 (<http://www.tal.univ-paris3.fr/trameur/#p9>).
15. Suivent des formes qui contiennent un sème relationnel, comme *famille*. Par ailleurs, on voit que la conjonction ne met pas prioritairement en relation des nominalisations dans le texte finalisé. Dans les jets textuels, ce sont également les désignations des personnes qui sont reliées en premier lieu.

RÉSUMÉS

Cet article se propose d'appréhender le segment du point de vue de la production, en analysant les séquences identifiées lors du processus d'écriture, c'est-à-dire les unités langagières produites de manière spontanée et ininterrompue (entre deux pauses). Ces unités, que nous appelons jets textuels, ont été enregistrées grâce à l'outil de suivi de rédaction *Inputlog*. Nous avons tenté de caractériser les jets textuels en les mettant en regard avec divers types d'unités segmentales de langage, qu'elles soient issues de la segmentation des textes finalisés, ou reconnues comme des segments préfabriqués propres à un système lexical ou syntaxique, c'est-à-dire des figements au sens large du terme. Nous nous sommes plus particulièrement penchés sur un parcours de segmentation parallèle, en production (jets textuels) et en réception (segments répétés), pour interroger quelques concepts théoriques et méthodologiques qui sous-tendent les démarches de segmentation : pré-construit, automatisme, mémorisation, figement, enracinement cognitif, fréquence, répétition.

This article aims at understanding the segment from the point of view of production, by analyzing the sequences identified during the writing process, *i.e.*, the language units spontaneously produced without interruption (between two pauses). These units, called bursts of writing, were recorded using *Inputlog*, a key-stroke logging tool. We have attempted to characterize the bursts of writing by comparing them with various types of language sequences, whether they are derived from the segmentation of finalized texts, or recognized as prefabricated segments specific to a lexical or syntactic system, like frozen units. In particular,

we have focused on a parallel segmentation process, in production (bursts of writing) and in reception (repeated segments), in order to question some theoretical and methodological concepts that underlie segmentation processes: pre-construction, automatism, memorization, cognitive entrenchment, frequency, repetition.

INDEX

Mots-clés : jets textuels, segments répétés, pré-construit, fréquence, automatisme

Keywords : bursts of writing, repeated segments, pre-construction, frequency, automatism

AUTEURS

GEORGETA CISLARU

Université Sorbonne Nouvelle

THIERRY OLIVE

CNRS – Université de Poitiers