

Titre : SEGMENTS et séquences textuelles : méthodologie et caractérisation

Coordinatrice : Marion Bendinelli (Université de Franche-Comté - ELLIADD/DTEPS)

Cette livraison de *Corpus* prendra pour objet d'étude les structures lexico-grammaticales plus ou moins figées (dorénavant SEGMENTS¹), désignées selon les courants théoriques, périodes et propriétés définitoires retenues (inventaire non exhaustif) : unités phraséologiques ou phraséologies, phrases² semi-préconstruites (Sinclair 1991), collocations (de nature lexicale ou grammaticale - Firth 1957, Halliday 1961 notamment), cadres collocationnels (Renouf & Sinclair 1991), colligations (de type lexical ou textuel - Hoey 2005), segments répétés (Salem 1986) et quasi-segments répétés (Bécue 1996), motifs (Longrée & Mellet 2012), unités lexicales étendues (Sinclair 2004), séquences ou patrons formulaires (Biber 2009), routines discursives (Née, Sitri, Veniard 2014), matrices lexicales (Anscombe 2011), patrons (Hunston & Francis 2000), constructions (Fillmore 1988 ; Bouveret & Legallois 2012 pour le français), constructions préformées (Schmale 2013)... Ces dénominations non-synonymiques ont en commun de postuler l'existence d'une **structure pré-formée** qui combine **deux ou plusieurs unités** (lexèmes, morphèmes, catégories grammaticales, contours prosodiques...) appartenant au **même niveau linguistique ou non** ; le sens de la structure peut varier le long d'un **continuum allant de la non-compositionnalité** (*i.e.*, le sens n'est pas déductible de ses constituants) **à la compositionnalité** et son **actualisation peut être sujette à variations lexicale et/ou morpho-syntaxique**. Le caractère pré-construit de ces SEGMENTS repose sur des **mécanismes inscrits en langue** (aux niveaux phonologique, prosodique, sémantique, morpho-syntaxique, énonciatif) **et/ou en discours** : enchaînements d'ordre discursif, pragmatique, voire structurations informationnelle (des énoncés précédents) et cohésive (du texte).

L'étude des SEGMENTS dépasse ainsi le cadre analytique traditionnellement constitué autour de la sémantique et de la syntaxe pour potentiellement toucher l'ensemble des niveaux d'analyse en linguistique. Cette perspective est déjà celle adoptée par Salah Mejri qui, dans le numéro 53 de la revue *Linx*, interroge la notion de degré de figement et conclut au fait que « [l]e figement couvre la totalité des champs de la linguistique. En tant que phénomène général et systématique, il impose une relecture de toute la syntaxe, la sémantique et l'analyse du discours » [*en ligne*]. Plus récemment, elle est au cœur du postulat de Dominique Legallois et Agnès Tutin pour qui l'inflation des objets étudiés par la phraséologie est tout à la fois significative « d'un tournant phraséologique de la linguistique et d'une extension du champ phraséologique » (2013 : 19). **C'est le regard de l'analyse de données outillée d'une part (lexicométrie, textométrie, logométrie, traitement automatique des langues, fouilles de données) et de la linguistique textuelle d'autre part que nous souhaitons ici favoriser et, si possible, confronter/mettre en dialogue, autour de deux problématiques : (1) comment détecter les SEGMENTS en corpus ? (2) ces SEGMENTS structurent-ils les séquences textuelles ?**

(1) Quel que soit le(s) ressort(s) explicatif(s) - situé(s) en *langue* ou en *discours* -, il apparaît que la définition des SEGMENTS suppose une attraction mutuelle entre deux ou plus unités, donc une co-apparition, fréquente, mesurable, dans un empan à définir. Un tel constat permet de convoquer les méthodes outillées issues de la statistiques textuelle (Lebart & Salem 1994) ou du traitement automatique des langues (Quiniou, Cellier, Charnois, Legallois 2012), appliquées à des textes réunis en corpus et enrichis de métadonnées, *a minima* un encodage

¹ Le recours aux petites majuscules indique que nous entendons le terme comme recouvrant l'ensemble des objets d'étude cités, faisant ainsi de lui une sorte d'hyperonyme de la classe.

² Au sens de « manière de dire ».

morpho-syntaxique et lexical ; on pensera ainsi pour le repérage des SEGMENTS aux concordanciers, segments répétés, calculs de co-occurrence, calculs de poly-cooccurrences (liste non-exhaustive). Toutefois, l'hétérogénéité des niveaux linguistiques sollicités dans les SEGMENTS, leur aptitude à accueillir une certaine dose de variation, complexifie sensiblement leur détection automatique.

(2) Plusieurs études montrent que les SEGMENTS sont à même de caractériser les genres discursifs, les registres de langue et/ou les auteurs : c'est ce que présente notamment Legallois (2012 : *en ligne*) autour du motif émergent du genre poétique « le N1 du N2 qui V dans ». C'est ici un autre niveau de caractérisation qui nous intéresse, celui de la structuration des séquences textuelles. Nous souhaitons ainsi interroger, dans un mouvement aussi bien descendant qu'ascendant, le rapport des segments au texte *via* les séquences textuelles, palier organisationnel du texte avec le genre, les relations de cohérence et les structures d'enchaînement (Legallois 2006).

Une séquence est, dans la définition qu'en donne Jean-Michel Adam,

à la fois une **unité constituée** (dont il faut alors décrire la structure interne et les *constituants*) et une **unité constituante** (dont il faut, dans le cas des textes comportant plusieurs séquences, décrire les modes d'enchaînement séquentiels : *insertion* et *dominante*). Comme unité constituante, **la séquence** est une composante de T (= texte) ; comme unité constituée, **la séquence** est composée de micro-propositions ». (1984 : 136)

Certains travaux (p. ex. Gledhill 2000 pour les articles scientifiques, Römer 2010 pour les critiques de livres) signalent, à des fins de caractérisation discursive ou générique³, la fonction narrative, introductive, conclusive, argumentative de SEGMENTS. De telles observations mènent à penser qu'une exploration plus poussée du rapport SEGMENT / séquence est souhaitable.

Une telle perspective soulève plusieurs questions : comment les SEGMENTS se répartissent-ils au sein des textes et séquences ? Ont-ils une place privilégiée à même de définir chaque séquence et de les opposer entre elles ? Cette dernière hypothèse semble confirmée par les travaux sur la colligation, qui, dans la définition de Michael Hoey, est associée à « the place in a sequence that a word prefers or avoids » (2005 : 43)⁴ ; on pense également aux éléments initiaux potentiellement définitoires du mode d'argumentation requis dans les articles scientifiques (Carter-Thomas 2014). D'autres expérimentations ont aussi montré qu'il est possible de cartographier la place de SEGMENTS *via* l'observation de leur ventilation au sein de textes préalablement segmentés (Römer 2010, Bendinelli 2014). De telles études laissent envisager la possibilité d'identifier les types de séquence constituant un texte en le segmentant selon la trame dessinée par les SEGMENTS préalablement détectés au sein de ce texte. Pour toutes ces questions, les méthodes outillées - statistique cooccurrentielle et poly-cooccurrentielle, représentation topologique des textes - peuvent, à nouveau (cf. point 1), représenter un appareillage pertinent pour le-la chercheur-e.

Il conviendrait donc d'**interroger la place (topologique) et le rôle (pragmatico-discursif) des SEGMENTS dans l'organisation et la structuration des séquences textuelles, en interne (entre propositions) et en externe (entre séquences et/ou sous-séquences)**. Toutes les séquences peuvent *a priori* être prises en considération : reprenant les travaux de Adam

³ À l'exception de Longrée & Mellet (2012) où la propriété structurante des *motifs* est partie intégrante de la définition de l'unité.

⁴ La notion de séquence n'est pas précisée par l'auteur.

(1992), complétés de ceux de Hyland (2005), Smith (2003) et Charaudeau (1992)⁵, on citera l'exemple des séquences narrative, descriptive, argumentative, explicative, dialogale, métadiscursive, instructionnelle, informationnelle. Il s'agirait de compléter la grammaire des séquences textuelles proposée par ces auteurs en enrichissant, au moyen des SEGMENTS, les propriétés formelles et sémantico-pragmatiques déjà identifiées. L'objectif est d'apporter une contribution aux études s'intéressant à l'analyse fonctionnelle des SEGMENTS, à la caractérisation générique et/ou stylistique des textes, à la recherche sur les discours spécialisés ainsi que, plus généralement, à la didactique des langues.

Les études devront porter sur des productions empiriques attestées, orales ou écrites, envisagées du point de vue de la situation de communication et de la pratique socio-historique qui président à leur existence. Ces textes (Rastier 2005), qui constituent l'unité minimale de l'analyse linguistique, ne prennent cependant sens qu'envisagés au sein d'un corpus (*ibid.*), faisant ainsi de ce dernier le « référentiel de l'analyse » (Bommier-Pincemin 1999). La construction du corpus devra répondre aux exigences posées par l'approche informatisée et statistique des données dont il rend compte (*ibid.*, Mayaffre 2002). Seront privilégiés les corpus représentatifs d'un genre discursif (débat, entretiens journalistiques, conférences de presse, spots publicitaires, article scientifique) et/ou d'une pratique socio-discursive (discours politique, discours scientifique, discours médiatique, discours d'expertise, etc.).

Des publications récentes en France, dans cette revue (*Corpus* n°11/2012, n°13/2014) et ailleurs (*Langages* n°189/2013), ont porté une attention toute particulière à la question de la cooccurrence, des éléments initiaux et au domaine de la phraséologie ; ce constat, associé à la quasi-concomitance de plusieurs appels à publication (Sitri & Tutin en 2014 autour des routines phraséologiques et des genres textuels ; Patard & Legallois en 2015 autour des constructions - inventaires, évaluations, critiques), fait des SEGMENTS un objet scientifique brûlant.

Références bibliographiques

- ADAM, J.-M., 1984, *Le Récit*, Paris : Presses Universitaires de France.
- _____, 1992, *Les textes: types et prototypes. Récit, description, argumentation, explication et dialogue*, Paris : Nathan.
- ANSCOMBRE J.-C., 2011, « Figement, idiomaticité et matrices lexicales », in ANSCOMBRE J.-C. & S. MEJRI (éds.), *Le figement linguistique: la parole entravée*, Paris : Ed. Champion 17-40.
- BECUE M., 1993, « Les quasi-segments pour une classification automatique des réponses ouvertes », *JADT 1993, Montpellier*. Paris, École nationale supérieure des télécommunications.
- BENDINELLI M., 2014, « Détection et analyse de séquences argumentatives. *Modaux et semi-modaux dans les débats présidentiels américains* », séminaire invitée du laboratoire FoReLL, Mars 2014, Université de Poitiers.
- BIBER D., 2009, « A corpus-driven approach to formulaic language: Multi-word patterns in speech and writing », *International Journal of Corpus Linguistics*, 14, 381-417.
- BOMMIER-PINCEMIN B., 1999, *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*. Thèse de Doctorat en Linguistique. Université Paris IV Sorbonne.
- BOUVERET M. & D. LEGALLOIS, 2012, *Constructions in French*, Amsterdam/Philadelphia : John Benjamin.

⁵ Carlota Smith et Patrick Charaudeau désignent respectivement ces unités *modes of discourse* et modes d'organisation du discours.

- CARTER-THOMAS S., 2014, « Valeurs et fonctions des éléments initiaux commentaires. Analyse contrastive d'un corpus d'articles de recherche en économie », *Corpus* [En ligne], 13, URL : <http://corpus.revues.org/2455>, consulté le 23 août 2015.
- CARTER-THOMAS S. & S. PRÉVOST, 2014, *Éléments initiaux dans la phrase : approches inter-genres et inter-langues*, *Corpus* [En ligne], 13, URL : <http://corpus.revues.org/2424>, consulté le 23 août 2015.
- CHARAUDEAU P., 1992, *Grammaire du sens et de l'expression*, Paris : Hachette Éducation
- FILLMORE C., 1988, « The mechanisms of construction grammar », *Berkeley Linguistic Society*, 14, 35-55.
- FIRTH J.R., 1957, *Papers in Linguistics (1934-1951)*, Oxford : Oxford University Press.
- GLEDHILL C., 2000, « The Discourse Functions of Collocation in Research Article Introduction », *English for Specific Purposes*, 19, 115-135.
- HALLIDAY M.A.K., 1961, « Categories of the theory of grammar », *Word*, 17(3), 241-92.
- HOEY M., 2005, *Lexical priming: A new theory of words and language*. Londres : Routledge.
- HUNSTON S. & G. FRANCIS, 2000, *Pattern Grammar : a Corpus-Driven Approach to the Lexical Grammar of English*, Amsterdam/Philadelphia : John Benjamin.
- HYLAND K., 2005, *Metadiscourse*, Londres : Continuum.
- LEGALLOIS D., 2006, « Quand le texte signale sa structure : la fonction textuelle des noms sous-spécifiés », *Corela* [En ligne], HS-5, URL : <http://corela.revues.org/1465>, consulté le 29 août 2015.
- _____, 2012, « La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? », *Corpus* [En ligne], 11, URL : <http://corpus.revues.org/2202>, consulté le 29 août 2015.
- LEGALLOIS D., & A. TUTIN, 2013, *Vers une extension du domaine de la phraséologie*, *Langages*, 189.
- LONGRÉE D. & S. MELLET, 2012, « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours », *Langages*, 189, 65-80.
- MAYAFFRE D., 2002, « Les corpus *réflexifs* : entre architextualité et hypertextualité », *Corpus*, 1, 51-69.
- MAYAFFRE D. & J.-M. VIPREY, 2011, La cooccurrence, du fait statistique au fait textuel, *Corpus*, 11.
- MEJRI S., 2005, « Figement absolu ou relatif : la notion de degré de figement », *Linx* [En ligne], 53, URL : <http://linx.revues.org/283>, DOI : 10.4000/linx.283, consulté le 22 août 2015.
- NÉE É., SITRI F., VENIARD M., 2014, « Pour une approche des routines discursives dans les écrits professionnels », Communication donnée lors du Congrès Mondial de Linguistique Française, Juillet 2014, Berlin, Allemagne, URL : <http://dx.doi.org/10.1051/shsconf/20140801195>, consulté le 22 août 2015.
- RÖMER U., 2010, « Establishing the Phraseological Profile of a Text Type: The Construction of Meaning in Academic Review Books », *English Text Construction*, 3(1), 95-119.
- RASTIER F., 2005, « Enjeux épistémologiques de la linguistique de corpus », in WILLIAMS Geoffrey (dir.), *La linguistique de Corpus*, Rennes : Presses Universitaires de Rennes, 31-46.
- RENOUF A. & J.M. SINCLAIR, 1991, « Collocational Frameworks in English », in AIJMER K. & B. ALTENBERG (eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, Londres : Longman, 128-143.
- SALEM A., 1986, « Segments répétés et analyse statistique des données textuelles », *Histoire & Mesure*, 1(2), 5-28, URL : http://web/revues/home/prescript/article/hism_0982-1783_1986_num_1_2_1518, DOI : 10.3406/hism.1986.1518, consulté le 22 août 2015.
- SCHMALE Günther, 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.
- SINCLAIR J.R., 1991, *Corpus, concordance, collocation: Describing English language*, Oxford : Oxford University Press.
- _____, 2004, *Trust the Text : Language Corpus and Discourse*, Londres : Routledge.
- SMITH C., 2003, *Modes of Discourse. The Local Structure of Texts*, Cambridge : Cambridge University Press.

TUTIN A., 2012, « Les collocations dans le champ sémantique des émotions: la régularité plutôt que l'idiosyncrasie », in APREJAN Y. *et al.* (éds.), *Meanings, texts and other exciting things : a festschrift to commemorate the 80th anniversary of Professor Igor Alexandrovic Mel'čuk*, Moscou : Jazyki slavanskoj kul'tury, 602-612.

Modalités de soumission

Merci d'adresser un **résumé long** de votre proposition d'article incluant une présentation de la problématique, des données, de la méthodologie ainsi qu'une courte bibliographie (2 pages maximum - format Word et Pdf) par courriel avant le **15 décembre 2015** à <marion.bendinelli@univ-fcomte.fr>.

Après acceptation, l'**article complet** devra être envoyé avant le **1^{er} avril 2016** et respecter la feuille de style de la revue *Corpus* (cf. site de la revue > Consignes aux auteurs).

Calendrier

15 septembre 2015 : diffusion de l'appel

15 décembre 2015 : réception des résumés longs accompagnés d'une lettre d'intention pour pré-sélection

31 décembre 2015 : notification de l'acceptation

1^{er} avril 2016 : réception des articles complets pré-sélectionnés

15 septembre 2016 : version finale pour édition

15 octobre 2016 : sortie du numéro 15 de *Corpus* (version papier)

15 avril 2017 : sortie du numéro 15 de *Corpus* (version électronique)

Contact

Marion Bendinelli

Université de Franche-Comté

EA 4661 ELLIADD/DTEPS

30-32 rue Mégevand

25000 Besançon

marion.bendinelli@univ-fcomte.fr