

## **Title: SEGMENTS and Textual Sequences: Methodology and Characterization**

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

This new issue of *Corpus* aims to study more or less fixed lexico-grammatical structures (from now on called SEGMENTS<sup>1</sup>) that, according to the theoretical branches, periods and definitory properties taken on, are named (non-exhaustive list): phraseological units or phraseologies, semi-preconstructed phrases<sup>2</sup> (Sinclair 1991), lexical or grammatical collocations (Firth 1957, Halliday 1961), collocational frameworks (Renouf & Sinclair 1991), lexical or textual colligations (Hoey 2005), repeated segments (Salem 1986) and repeated quasi-segments (Bécue 1996), motifs (Longrée & Mellet 2012), extended lexical units (Sinclair 2004), formulaic sequences (Biber 2009), discursive routines (Née, Sitri, Veniard 2014), lexical matrix (Anscombe 2011), patterns (Hunston & Francis 2000), constructions (Fillmore 1988 ; Bouveret & Legallois 2012 for French), prefabricated constructions (Schmale 2013), *etc.*

These non-synonymic objects share some characteristics: all postulate the existence of a **prefabricated structure** combining **two or more units** (lexemes, morphemes, parts of speech, prosodic contours, *etc.*) which may or may not be located on **the same linguistic level**; the meaning of the structure can vary along a **continuum, from non-compositionality** (*i.e.*, the meaning cannot be deduced from its constituents) to **compositionality**, and its **realization can be subject to lexical and/or morpho-syntactic variations**. The prefabricated nature of the SEGMENTS may rest on **linguistic** mechanisms (phonology, prosody, semantics, morpho-syntax, enunciation) **and/or discursive** phenomena: discursive moves, pragmatic sequences, informational structure of successive utterances, cohesive structure of texts.

Studying segments implies going beyond the traditional analytic scope of semantics and syntax in order to potentially achieve all linguistic levels. Salah Mejri has already adopted this point of view and he examines, in the 53rd issue of *Linx*, the degree of fixedness. He concludes that « fixedness covers the whole field of linguistics. As a general and invariable phenomenon, it requires checking the whole syntax, semantics and discourse analysis » [*online* - our translation]. More recently, the degree of fixedness is the premise of Dominique Legallois and Agnès Tutin who consider that the increasing number of objects studied in phraseology reveals « a phraseological turning point of linguistics and an expansion of the field of phraseology » (2013 : 19 - our translation). **We would like to favour in this issue two research directions, and if possible confront them: (1) are computer-assisted methods (lexicometry, textometry, logometry, natural language processing, data mining) able to detect SEGMENTS in a corpus, and how? (2) which new perspective(s) does text linguistics offer to the analysis of SEGMENTS, with the subsequent question: do SEGMENTS structure textual sequences?**

(1) Whether SEGMENTS are explained in linguistic and/or discursive terms, their realization rests on the attraction of two or more units, *i.e.*, a regular and measurable co-appearance within a given span (whose size is to be defined). Such an analysis makes it possible to use computer-assisted methods, developed in the field of textual statistics (Lebart & Salem 1994) or natural language processing (Quiniou, Cellier, Charnois, Legallois 2012); these are applied on textual data that are compiled in a corpus and documented with metadata such as morpho-

---

<sup>1</sup> We use small caps for this word because it covers all the mentioned research objects. Therefore, it becomes a kind of hyperonym of the category.

<sup>2</sup> With the meaning of « manner of speaking ».

syntactical and lexical tags. Thus, the use of concordancers, repeated segments, co-occurrence and poly-cooccurrence networks (non-exhaustive list) may be relevant to automatically detect SEGMENTS; however, such a task is complicated by the heterogeneous nature of the linguistic levels involved in the realization of SEGMENTS as well as by their intrinsic ability to accept variations.

(2) Many studies show that SEGMENTS can be characteristic of discourse genres, registers and/or authors: it is what Legallois (2012: *online*) acknowledges with « le N1 du N2 qui V dans » considered as an emerging motif of the poetic genre. Here, we are interested in another level of characterization: the structure of textual sequences. We aim to examine, with a top-down and bottom-up approach, the links between SEGMENTS and text through an intermediary component, textual sequences: these help organize texts, together with discourse genre, coherence and cohesive moves (Legallois 2006).

A sequence is, according to Jean-Michel Adam,

a unit that is both **constituted** (for which we have to describe the internal structure and the components) and **constituting** (for which we have to describe the sequence structures - *insertion* and *dominant feature* - in the case of texts including many sequences). As a constituting unit, **the sequence** is an component of T (= text) ; as a constituted unit, **the sequence** is composed of micro-propositions. (1984: 136 - our translation)

Previous works (see Gledhill 2000 for scientific papers, Römer 2010 for book reviews) highlight, for purposes of discursive or generic characterization<sup>3</sup>, the narrative, introductory, conclusive, argumentative function of SEGMENTS. Such observations call for a more advanced study of the link between SEGMENT and textual sequences.

This possibility raises many questions: how are SEGMENTS distributed in texts and sequences? Do they favor a position which would help define each sequence and confront one with another? The latter hypothesis appears to be confirmed by works studying (i) the phenomenon of colligation, which is connected with « the place in a sequence that a word prefers or avoids » (Hoey 2005 : 43)<sup>4</sup>; (ii) the initial elements which potentially define the type of argumentation required in scientific papers (Carter-Thomas 2014). Other experiments have also proved that mapping the position of segments is possible by observing their distribution in previously segmented texts (Römer 2010, Bendinelli 2014): such works suggest that it may be possible to determine the sequence(s) that shape(s) a given text on the basis of the nature and distribution of previously detected SEGMENTS. Again (see the first point), for all these questions, computer-assisted methods – co-occurrence and poly-cooccurrence statistics, topological representations of texts for example – can be relevant tools for researchers.

On such basis, papers are expected to **examine the (topological) position and the (pragmatic-discursive) role of SEGMENTS in the organization and structure of textual sequences, both internally (between clauses) and externally (between sequences and/or sub-sequences)**. All types of sequence can be in principle considered: based on Adam's works (1992) as well as Hyland's (2005), Smith's (2003) and Charaudeau's works (1992)<sup>5</sup>,

---

<sup>3</sup> Except for Longrée & Mellet (2012) who think the structuring attribute of motifs is an integral part of the unit definition.

<sup>4</sup> The notion of sequence is not defined by the author.

<sup>5</sup> Carlota Smith and Patrick Charaudeau respectively name these units “modes of discourse” and “modes of discourse organization”.

we can mention for instance, narrative, descriptive, argumentative, explanatory, dialogal, metadiscursive, instructional and informational sequences. An objective would be to complete the grammar of textual sequences suggested by these authors and expand, thanks to SEGMENTS, the formal and semantico-pragmatic attributes which have already been identified. The intent is therefore to contribute to studies concerned with the functional analysis of segments, the generic and/or stylistic characterization of texts, as well as with specialized discourses and more generally didactics.

Propositions of papers may focus on oral or written empirical productions defined in terms of a situation of communication and a socio-historical practice. Such texts (Rastier 2005) constitute the minimal unit of linguistic analysis; they can make sense only if they are considered within a corpus (*ibid.*) which becomes a « frame of reference for the analysis » (Bommier-Pincemin 1999 - our translation). The corpus must be elaborated in accordance with the requirements implied by the use of a computer-assisted approach and statistical processing (*ibid.*, Mayaffre 2002). Corpora that are representative of a particular discourse genre (debates, journalistic interviews, press conferences, commercials, scientific papers) and/or of a given socio-discursive practice (political discourse, scientific discourse, media discourse, expert discourse, *etc.*) are favored.

Co-occurrence, initial elements and phraseology have been particularly studied in recent French papers published in this journal (*Corpus* n°11/2012, n°13/2014) as well as in *Langages* (n°189/2013). This observation combined with several quasi-simultaneous calls for papers (Sitri & Tutin in 2014 about linguistic routines and text types; Patard & Legallois in 2015 about structures – lists, assessments, criticisms) make SEGMENTS a burning research subject.

### **Bibliographical references**

- 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, idiomatité 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.

- CARTER-THOMAS S., 2014, « Valeurs et fonctions des éléments initiaux commentaires. Analyse contrastive d'un corpus d'articles de recherche en économie », *Corpus* [On line], 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* [On line], 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*, London: 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*, London: Continuum.
- LEGALLOIS D., 2006, « Quand le texte signale sa structure : la fonction textuelle des noms sous-spécifiés », *Corela* [On line], 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* [On line], 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* [On line], 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 Ute, 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*, London: 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](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*, London: 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*, Moscow: Jazyki slavanskoj kul'tury, 602-612.

---

### **Submission modalities**

**Long abstracts** (two pages) should be sent as Word and PDF documents attached to the message before **December 15<sup>th</sup> 2015** at [marion.bendinelli@univ-fcomte.fr](mailto:marion.bendinelli@univ-fcomte.fr). They should include a presentation of the project, the data and the methodology as well as a short bibliography.

Selected submissions will be sent in their **full version** by **April 1<sup>st</sup> 2016** in accordance with *Corpus* style sheet (see website > Consignes aux auteurs).

### **Submission deadline**

September 15<sup>th</sup> 2015 : call for papers

**December 15<sup>th</sup> 2015 : submission of long abstracts for selection together with a letter of intent**

December 31<sup>st</sup> 2015: notification of acceptance

**April 1<sup>st</sup> 2016 : submission of articles (full versions)**

September 15<sup>th</sup> 2016 : final version for edition purpose

October 15<sup>th</sup> 2016 : issue of *Corpus* n°15 (paper version)

April 15<sup>th</sup> 2017 : issue of *Corpus* n°15 (online version)

### **Contact**

Marion Bendinelli

Université de Franche-Comté

EA 4661 ELLIADD/DTEPS

30-32 rue Mégevand

25000 Besançon

[marion.bendinelli@univ-fcomte.fr](mailto:marion.bendinelli@univ-fcomte.fr)