



HAL
open science

Multidisciplinary Perspectives on Signalling Text Organisation

Lydia-Mai Ho-Dac, Julie Lemarié, Marie-Paule Péry-Woodley, Marianne Vergez-Couret

► **To cite this version:**

Lydia-Mai Ho-Dac, Julie Lemarié, Marie-Paule Péry-Woodley, Marianne Vergez-Couret. Multidisciplinary Perspectives on Signalling Text Organisation. *Discours - Revue de linguistique, psycholinguistique et informatique*, 2012, 10, <http://discours.revues.org/8598>. hal-00730361

HAL Id: hal-00730361

<https://hal.science/hal-00730361v1>

Submitted on 9 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Lydia-Mai Ho-Dac, Julie Lemarié, Marie-Paule Péry-Woodley et
Marianne Vergez-Couret

Multidisciplinary Perspectives on Signalling Text Organisation: Introduction to the Special Issue

Perspectives multidisciplinaires sur la signalisation
de l'organisation textuelle : introduction au numéro
thématique

Avertissement

Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur.

Les œuvres figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage soit personnel, soit scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document.

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

revues.org

Revues.org est un portail de revues en sciences humaines et sociales développé par le Cléo, Centre pour l'édition électronique ouverte (CNRS, EHESS, UP, UAPV).

Référence électronique

Lydia-Mai Ho-Dac, Julie Lemarié, Marie-Paule Péry-Woodley et Marianne Vergez-Couret, « Multidisciplinary Perspectives on Signalling Text Organisation: Introduction to the Special Issue », *Discours* [En ligne], 10 | 2012, mis en ligne le 16 juillet 2012, consulté le 09 septembre 2012. URL : <http://discours.revues.org/8598> ; DOI : 10.4000/discours.8598

Éditeur : Laboratoire LATTICE, UMR 8094 ENS/CNRS

<http://discours.revues.org>

<http://www.revues.org>

Document accessible en ligne sur : <http://discours.revues.org/8598>

Ce document PDF a été généré par la revue.

© Discours - Presses universitaires de Caen



Revue de linguistique, psycholinguistique et informatique

<http://discours.revues.org/>

Multidisciplinary Perspectives on Signalling Text Organisation: Introduction to the Special Issue

Perspectives multidisciplinaires sur la signalisation de
l'organisation textuelle : introduction au numéro thématique

Lydia-Mai Ho-Dac

CLLE-ERSS, CNRS & Université de Toulouse (UTM)

Julie Lemarié

CLLE-LTC, CNRS & Université de Toulouse (UTM)

Marie-Paule Péry-Woodley

CLLE-ERSS, CNRS & Université de Toulouse (UTM)

Marianne Vergez-Couret

LPL, CNRS & Université d'Aix-en-Provence – Université polytechnique de Hong Kong

.....
Lydia-Mai Ho-Dac, Julie Lemarié, Marie-Paule Péry-Woodley et Marianne Vergez-Couret, « Multidisciplinary Perspectives on Signalling Text Organisation: Introduction to the Special Issue », *Discours* [En ligne], 10 | 2012, mis en ligne le 16 juillet 2012.

.....
URL : <http://discours.revues.org/8598>

.....
Titre du numéro : *Multidisciplinary Perspectives on Signalling Text Organisation*

Coordination : Shirley Carter-Thomas & Frédéric Landragin

revues.org
CENTRE POUR L'ÉDITION ÉLECTRONIQUE OUVERTE
CENTRE FOR OPEN ELECTRONIC PUBLISHING

 discours

 Presses
universitaires
de Caen

Multidisciplinary Perspectives on Signalling Text Organisation: Introduction to the Special Issue

Perspectives multidisciplinaires sur la signalisation de
l'organisation textuelle : introduction au numéro thématique

Lydia-Mai Ho-Dac

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

Julie Lemarié

CLLE-LTC

CNRS & Université de Toulouse (UTM)

Marie-Paule Péry-Woodley

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

Marianne Vergez-Couret

LPL

CNRS & Université d'Aix-en-Provence – Université polytechnique de Hong Kong

Le lecteur trouvera une version française de cette introduction ci-dessous.

- 1 Texts are organised wholes. Understanding a text entails constructing a representation of its organisation. Several research domains, with different assumptions and objectives, have taken an interest in the devices which seem to help readers in this process. As a consequence, research concerned with the signalling of text organisation is far from constituting a unified field. The notion of signal itself may be associated with different key concepts according to discipline and theoretical underpinning: document structure, discourse organisation, layout structure, text architecture, etc. As far as function is concerned, signals may be seen as discourse construction devices, as metadiscourse, as reading or processing instructions, as traces of the writer's cognitive processes, as cues revealing the author's intentions... This special issue, which follows on from the MAD 2010 workshop in Moissac (<http://w3.workshop-mad2010.univ-tlse2.fr>), is a first step towards the development of multidisciplinary approaches: by bringing together perspectives from different disciplines interested in the signalling of text organisation —linguists, computational linguists, psycholinguists, educational/cognitive psychologists, document designers— we aim to contribute to mutual understanding

and cross-fertilization. We urge readers to resist the temptation to stay smugly within their own field, and invite them to find out what research in other disciplines brings up in relation to text organisation signals.

2 Though this issue is modestly multidisciplinary in the sense that it mostly brings together research from different fields, it is also more ambitiously so in the first two papers, which present approaches associating several disciplines: cognitive psychology and linguistics for the first, information design (and architecture...) and linguistics for the second.

3 In their paper on “Understanding How Headings Influence Text Processing”, **Julie Lemarié, Robert F. Lorch** and **Marie-Paule Péry-Woodley** show that the lack of integration of linguistic and psychological research concerning signals in general and headings in particular leads to blind spots; they present a general theoretical framework, SARA, that provides a linguistically-based approach to analyse signalling devices and generates psychological predictions on their effects on text processing by the reader. They illustrate how this integrative framework may be used to investigate how headings influence text processing.

4 In their contribution, “Towards a Pattern Language Approach to Document Description”, **Robert Waller, Judy Delin** and **Martin Thomas** draw connections between information design, discourse research and corpus linguistics. They introduce the concept of pattern languages and pattern libraries, originating in architecture and widely used in software engineering and interaction design, as the basis for a functional description of the graphic structure of discourse. In information design, patterns are described as clusters of graphical and linguistic features that recur as regular solutions to particular design problems in particular contexts. The authors argue for a corpus approach with systematic multiple tagging of patterns in different genres and describe their own work in this direction. Among the features to be tagged, they stress the linguistic component of patterns, which they see as interacting closely with the graphical components.

5 This view of signals in terms of recurrent clusters of graphical and linguistic features is also defended by **Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle** and **Ludovic Tanguy** in “An Empirical Approach to the Signalling of Enumerative Structures”. They report on a large-scale corpus-based study of enumerative structures made possible by the recent availability of a corpus hand-annotated with discourse structures and cues. Taking the annotated structures as their starting point, and through the use of statistical methods, they propose an analysis of signalling as complex configurations of cues. On this empirical basis, they question the notion of textual metadiscourse as separate from propositional content, stressing instead the interplay of the three Hallidayan metafunctions: ideational, textual and interpersonal.

6 The next paper, “Genre Realised in Theme: The Case of News Reports and Commentaries”, by **Julia Lavid, Jorge Arús** and **Lara Moratón**, is also rooted in

the systemic functional perspective. The authors propose a comparative analysis of the thematic features appearing in “thematic heads” i.e. elements occurring in the preverbal zone at the beginning of a sentence. Their observations are based on a systematic human annotation of the semantic properties of these thematic heads in news reports vs. news commentaries. Their results show that each genre selects different “thematic participants” (e.g. the *Sayer* in reports) and uses different “thematic realisations” (e.g. concrete nouns in reports vs. abstract nouns in commentaries). Such semantic features are presented by the authors as features which, in a broad sense, “signal” these genres.

- 7 The next two papers are contributions to a bottom-up approach to text organisation using the framework developed within Segmented Discourse Representation Theory (SDRT). Bottom-up approaches to discourse structure attempt to give a representation of the structure through discourse relations between elementary segments. Relations may be signalled in text by means of so called discourse markers.
- 8 The article by **Charlotte Roze**, **Laurence Danlos** and **Philippe Muller**, “LEXCONN: A French Lexicon of Discourse Connectives”, contributes to the theme of the special issue by presenting a lexicon of 328 French discourse connectives or discourse markers along with their syntactic categories and the discourse relations they may express within the SDRT set of relations. The methodology consists firstly in applying various criteria, including syntactic, semantic and discourse criteria, to the identification of discourse connectives, and secondly in observing their context to associate each of them with one or several discourse relations it may express. Finally, the authors reveal some problem cases when no discourse relation in SDRT can describe the discourse contribution of the connective.
- 9 In their article, “Signalling *Elaboration*: Combining French Gerund Clauses with Lexical Cohesion Cues”, **Marianne Vergez-Couret** and **Clémentine Adam** tackle the question of the signalling of text organisation in cases where there are no obvious discourse markers, as it is usually the case for the *Elaboration* relation. The authors focus on this relation and on its automated detection. In the absence of specialised discourse markers, they investigate the presence of lexical cohesion links involved in the interpretation of *Elaboration*, and how to bring them to light using a lexical resource based on distributional similarity in combination with an ambiguous marker of *Elaboration*, the French gerund construction.
- 10 The last paper describes a study which associates a linguistic perspective on signalling with a psycholinguistic experimental approach: “Children’s Use of Comparative Text Signals: The Relationship between Age and Comprehension Ability” by **Bonnie J.F. Meyer**, **Melissa N. Ray** and **Wendy Middlemiss**. The authors offer a very original developmental perspective on the topic: they investigate how children and young adults acquire and use comparative signalling devices. Their results show different patterns of acquisition as a function of young readers’ comprehension skills.

- 11 Les textes étant des ensembles organisés, comprendre un texte passe par la construction d'une représentation de cette organisation. Différentes disciplines s'intéressent aux dispositifs qui guident le lecteur dans ce processus de construction, avec des hypothèses, des méthodes et des objectifs très variés. Il en résulte que les différents travaux de recherche relatifs à la signalisation de l'organisation des textes sont loin de constituer un champ unifié. La notion même de «signalisation» est associée à différents concepts selon les disciplines et les théories adoptées: structure de document, organisation discursive, structure visuelle, architecture textuelle, etc. Du point de vue fonctionnel, les signaux (*i. e.* les éléments signalant l'organisation des textes) peuvent être vus comme des dispositifs participant à la construction du discours, des indications métadiscursives, des instructions de lecture et de traitement du texte, des traces reflétant les processus de rédaction, des indices révélant les intentions de l'auteur... Ce numéro thématique consacré à la «signalisation de l'organisation des textes», qui fait suite au colloque MAD organisé à Moissac en mars 2010 (<http://w3.workshop-mad2010.univ-tlse2.fr>), constitue une première étape vers le développement d'approches multidisciplinaires: en rassemblant des perspectives issues de différentes disciplines intéressées par la signalisation de l'organisation du texte – linguistique, TAL (Traitement Automatique du Langage), psycholinguistique, psychologie cognitive, «document design» –, nous cherchons à favoriser la compréhension mutuelle et la fertilisation croisée. Nous encourageons ainsi le lecteur à résister à la tentation de rester confortablement dans son champ d'étude et l'invitons à découvrir l'apport d'autres disciplines à la question de la «signalisation de l'organisation des textes».
- 12 Bien que l'ambition multidisciplinaire de ce numéro spécial réside modestement dans le regroupement de travaux issus de différents domaines, les deux premiers articles associent plus audacieusement différentes disciplines: psychologie cognitive et linguistique pour le premier, linguistique et «document design» (et architecture...) pour le second.
- 13 Dans leur article intitulé «Understanding How Headings Influence Text Processing», **Julie Lemarié**, **Robert F. Lorch** et **Marie-Paule Péry-Woodley** montrent comment le manque d'intégration des recherches réalisées en linguistique et en psychologie à propos des signaux en général et des titres en particulier a conduit à la formation de points aveugles. Les auteurs présentent le cadre d'analyse SARA, qui propose de partir d'une description linguistique des dispositifs de signalisation pour prédire les effets de ces dispositifs sur les processus de compréhension. Ils démontrent comment ce nouveau cadre intégratif et pluridisciplinaire peut être utilisé pour étudier les titres et leurs effets.
- 14 Dans l'article «Towards a Pattern Language Approach to Document Description», **Robert Waller**, **Judy Delin** et **Martin Thomas** établissent des liens entre design de l'information, linguistique du discours et linguistique de corpus. Dans le but de décrire d'un point de vue fonctionnel la structure graphique des documents, les auteurs empruntent les notions de «langage de patrons» et de «bibliothèques de patrons» issues de l'architecture et déjà utilisées en conception logicielle et en design

numérique. Dans le domaine du design de l'information, les «patrons» sont des ensembles de traits graphiques et linguistiques utilisés pour répondre à un problème de design spécifique dans un contexte applicatif particulier. Les auteurs proposent une approche en corpus basée sur l'annotation systématique de ces patrons dans différents genres, en insistant particulièrement sur l'importance des traits de nature linguistique et leurs interactions avec les traits de nature graphique.

15 Cette idée d'interaction entre traits visuels et traits linguistiques est également défendue par **Lydia-Mai Ho-Dac**, **Cécile Fabre**, **Marie-Paule Péry-Woodley**, **Josette Rebeyrolle** et **Ludovic Tanguy** dans leur article «An Empirical Approach to the Signalling of Enumerative Structures». Cet article décrit une étude en corpus à grande échelle des structures énumératives sur la base d'une campagne d'annotation de ces structures et de leur signalisation. En partant des données annotées et en appliquant des méthodes statistiques de fouille, cette étude définit la signalisation en termes de configurations d'indices complexes. Par cette démarche empirique, les auteurs remettent en question la traditionnelle séparation entre contenu propositionnel d'une part et métadiscours d'autre part, et proposent une approche qui met l'accent sur la complémentarité entre les trois composantes textuelles définies en systémique fonctionnelle : composante idéationnelle, textuelle et interpersonnelle.

16 L'article suivant, «Genre Realised in Theme: The Case of News Reports and Commentaries», de **Julia Lavid**, **Jorge Arús** et **Lara Moratón**, est également ancré dans le modèle de la systémique fonctionnelle. Les auteurs présentent une analyse contrastive des éléments apparaissant en position «Thème», c'est-à-dire dans la zone préverbale en initiale de phrase. Leur étude se base sur une annotation manuelle systématique des propriétés sémantiques de ces «Thèmes» dans la presse en distinguant reportages et commentaires. Leurs résultats montrent que chaque genre sélectionne des «participants thématiques» différents et des «formes thématiques» différentes (noms concrets dans les reportages *vs* noms abstraits dans les commentaires). Ces traits sont vus par les auteurs comme des signaux spécifiques à ces différents genres.

17 Les deux articles suivants se situent dans l'approche ascendante du discours, ici représentée par le modèle de la SDRT (*Segmented Discourse Representation Theory*). Les approches ascendantes cherchent à représenter le discours en proposant un ensemble de relations de discours s'établissant entre unités de discours élémentaires. Ces relations peuvent être signalées à la surface du texte par ce qu'on appelle des «marqueurs discursifs».

18 Dans «LEXCONN: A French Lexicon of Discourse Connectives», **Charlotte Roze**, **Laurence Danlos** et **Philippe Muller** présentent les résultats du projet LEXCONN: un lexique de 328 connecteurs/marqueurs discursifs pour le français, dans lequel chaque marqueur est associé à sa fonction syntaxique et à la (ou aux) relation(s) de discours qu'il peut signaler parmi les relations proposées dans le cadre de la SDRT. La constitution d'un tel lexique repose tout d'abord sur l'identification des connecteurs en appliquant différents critères sémantiques, syntaxiques et discursifs. L'étape suivante consiste à observer ces connecteurs en contexte afin de leur associer

la ou les relations de discours qu'ils peuvent exprimer. Enfin, les auteurs discutent quelques cas problématiques pour lesquels aucune relation proposée par la SDRT ne convient à la situation discursive observée.

19 Dans leur article «Signalling *Elaboration*: Combining French Gerund Clauses with Lexical Cohesion Cues», **Marianne Vergez-Couret** et **Clémentine Adam** s'intéressent au cas où aucun marqueur lexical ne signale une relation, ce qui est souvent le cas pour la relation d'*Élaboration*. Dans l'optique de développer un outil de détection automatique de cette dernière, les auteures examinent la présence de liens de cohésion lexicale et leurs implications dans l'interprétation de la relation d'*Élaboration*, ainsi que la façon de les repérer automatiquement. Elles proposent de rendre compte de l'intérêt des indices de cohésion lexicale par l'utilisation croisée d'une ressource distributionnelle du français et d'un marqueur ambigu de la relation d'*Élaboration*: les constructions gérondives.

20 Le dernier article, «Children's Use of Comparative Text Signals: The Relationship between Age and Comprehension Ability», proposé par **Bonnie J.F. Meyer**, **Melissa N. Ray** et **Wendy Middlemiss**, rapporte une étude qui associe une perspective linguistique de la signalisation avec une méthode expérimentale psycholinguistique. Les auteures offrent une approche développementale du thème très originale: elles étudient comment les enfants et les jeunes adultes acquièrent et utilisent les marqueurs de comparaison. Leurs résultats montrent que les schémas d'acquisition varient en fonction des habiletés de lecture des enfants et des jeunes adultes.