



HAL
open science

Structures spatio-linguistiques du texte : traitements formels et cognitifs

Claudine Garcia-Debanc

► **To cite this version:**

Claudine Garcia-Debanc. Structures spatio-linguistiques du texte : traitements formels et cognitifs. 2005. hal-00003508

HAL Id: hal-00003508

<https://hal.science/hal-00003508>

Preprint submitted on 20 Jan 2005

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.

Structures spatio-linguistiques du texte : traitements formels et cognitifs

Responsable scientifique : Claudine GARCIA-DEBANC

Claudine GARCIA-DEBANC

Laboratoire Jacques Lordat, Université de Toulouse le Mirail
Maison de la Recherche, Université de Toulouse le Mirail
5 allée Antonio Machado
31058 TOULOUSE Cedex
Tél./Fax : 05 62 57 04 78
E-mail : claudine.garcia-debanc@toulouse.iufm.fr

Sous-thèmes dont relève ce projet :

Étude de certains
types d'espaces organisés :
espace du texte

Équipes partenaires

- IRIT, Institut de recherche en Informatique de Toulouse, Université Paul Sabatier, 118, Route de Narbonne, 31062 TOULOUSE CEDEX
- ERSS, Équipe de Recherche en Syntaxe et Sémantique, Université de Toulouse le Mirail, 5 allées Antonio Machado, 31058 TOULOUSE Cedex.
- LIMSI-CNRS, BP 133, ORSAY
- Laboratoire de Psychologie Expérimentale et Quantitative, Université de Nice Sophia Antipolis, 24 Avenue des Diabes Bleus, 06000 NICE

Résumé signalétique

La page, qu'elle soit écran ou papier, est un objet d'étude privilégié pour aborder la cognition spatiale. En effet, elle constitue un espace circonscrit, construit et régi par des conventions. Par « structures spatio-linguistiques du texte », nous entendons l'inscription d'instructions de lecture sous une forme lexico-syntaxique, typographique (tirets, gras, italiques, polices, etc.) et/ou dispositionnelle (blanc, indentation, alinéa, etc.). Nous proposons une approche organisée autour de trois axes mettant en œuvre respectivement une modélisation linguistique, une formalisation informatique et une expérimentation psycholinguistique. Nos principaux objectifs ont été (a) d'éclairer la relation entre la composante lexico-syntaxique du texte et son inscription visuelle dans la page, (b) de construire des outils d'extraction de l'information et de contrôle d'images de texte (c) d'étudier l'incidence de la structure visuelle du texte sur sa compréhension/mémorisation chez l'enfant et chez l'adulte, ainsi que le rôle spécifique des propriétés visuelles sur les stratégies de lecture. Les volets cognitif, linguistique et informatique s'enracinent dans le modèle de l'Architecture Textuelle, lui-même fondé sur les notions d'actes de langage et de mise en forme matérielle comme trace du métalangage textuel. L'analyse de corpus d'énumérations a permis de déduire des modèles descriptifs des objets textuels étudiés : énumérations et titres. Des avancées théoriques ont été réalisées dans le champ de la composition des modèles à travers la mise en relation du modèle d'Architecture du Texte et d'autres théories du texte comme la RST, qui ne prennent pas en compte les aspects visuels. D'un point de vue informatique, ont été élaborés un moissonneur de pages Web, un analyseur de titres et énumérations, un analyseur des initialisateurs (ou amorces) un outil de représentation et manipulation d'images de texte, des patrons de fouilles. D'un point de vue psycholinguistique, les expérimentations réalisées auprès d'adultes et d'enfants de 9 à 11 ans ont fait apparaître une influence de la mise en forme matérielle des textes sur le traitement en temps réel de structures énumératives et sur la compréhension/mémorisation d'informations contenues dans des textes de vulgarisation scientifique et des règles de jeu. Ont été également mis en évidence les effets de la nature syntaxique de l'amorce d'une structure énumérative et de la mise en forme matérielle des items de l'énumération sur sa mémorisation et son rappel.

Mots-clés : Mémoire spatiale • Lecture • Modèles de texte • Architecture de texte • Marqueurs d'organisation textuelle

Nombre de participants : Informatique : 5 ; Sciences du langage : 5 ; Psychologie : 2.

Nombre total d'hommes-mois : 52

Rappel des enjeux et objectifs fixés à l'origine

Comme l'indique le projet, la proposition présentée est centrée sur les structures spécifiquement textuelles (i.e. : relatives à des énoncés inscrits), et parmi celle-ci les structures visuelles et leurs relations avec d'autres types de structures textuelles. Elle repose sur un ensemble d'hypothèses de nature pragmatique et linguistique, psychologique et psycholinguistique, et enfin logico-informatique, qui visent à préciser le statut cognitif de la spatialité du texte, et plus particulièrement à :

- élucider les relations entre composante directement significative et composante d'efficacité communicationnelle et entre composante symbolique (morphologie) et composante dispositionnelle ;
- modéliser cette part de la face visuelle du langage dans ses relations avec le contenu lexical-syntaxique ;
- analyser l'effet des structures visuelles et textuelles sur la compréhension/mémorisation de textes par des adultes et par des enfants
- préciser le rôle des propriétés visuelles des textes dans les stratégies de lecture développées par les sujets en vue de construire une représentation interne du contenu ;
- vérifier les dimensions du renouvellement qu'apporte la modalité numérisée du texte, et ses formes de présentation et d'accessibilité (par ex. sur Internet).

Notre projet était plus précisément d'étudier les modalités d'organisation spatiale de certains objets textuels suivant un point de vue :

- psycholinguistique : impact de la forme visuelle du texte sur sa compréhension et sa mémorisation ;
- logico-linguistique et informatique : (1) caractérisation

- structurale et fonctionnelle des objets textuels ;
- 2) représentation et manipulation de ces objets textuels ;
- 3) application à l'extraction d'information.

Les objets textuels mentionnés dans le projet initial étaient les énumérations, les définitions et les titres. Ce choix est déterminé par la conjonction de considérations théoriques et applicatives : ces objets textuels (a) sont des structures textuelles dont l'identité tient beaucoup à l'organisation spatiale et (b) permettent de tester la validité de méthodes d'extraction automatique d'informations lexicales. Ils se situent souvent à la limite de la frontière qui signale les ressources spécifiques de la spatialisation.

Du point de vue logico-linguistique, le premier objectif est de répertorier les marques de signalisation de ces objets textuels. Ces marques correspondent aux traces laissées par l'effacement du métalangage associé à l'acte textuel de définir, d'énumérer ou de titrer. Elles peuvent être de nature lexicales, syntaxiques, typographiques et dispositionnelles.

D'un point de vue psycholinguistique, le projet se proposait d'analyser les effets de la disposition spatiale sur le traitement cognitif des objets textuels. Il s'agit de voir dans quelle mesure la disposition spatiale du texte dans la page peut favoriser la compréhension/mémorisation d'informations. L'étude se proposait de répondre notamment aux questions suivantes : Quelle est la version la plus efficace pour la compréhension/mémorisation d'informations ? A partir de quel âge les enfants parviennent-ils à utiliser efficacement les indices liés à la mise en forme matérielle des textes pour comprendre et mémoriser des informations ?

Résumé des résultats effectivement atteints

Les travaux effectués par les différentes équipes (IRIT-ERSS-LIMSI-LPEQ-LORDAT) dans le cadre de ce projet ont pour point commun l'étude de structures textuelles particulières exploitant pleinement les possibilités offertes par la spatialisation du langage et de leurs effets en terme de traitement cognitif. La démarche est sous tendue par le cadre théorique du modèle d'architecture textuelle.

D'un point de vue linguistique, nous considérons de façon conjointe des marques typo-dispositionnelles et lexico-syntaxiques comme marqueurs linguistiques de structures textuelles. À partir de cette hypothèse nous avons :

1. constitué des corpus d'énumérations : de manière empirique (IRIT), systématique (ERSS) ou automatique via un outil informatique (LIMSI) ;
2. analysé ces données (suivants les 3 types de marques) et
3. déduit des modèles descriptifs des objets étudiés : énumérations (ERSS) et titres (IRIT)

D'un point de vue logico-linguistique, les travaux ont permis de :

1. dresser un modèle de l'amorce (ou initialiseur) basée sur la

notion de prédicats sémantiques (LIMSI) ;

2. proposer un *mode de composition* pour des modèles de structuration du texte autorisant la prise en compte de phénomènes complexes comme les énumérations (IRIT) ;
3. donner une *typologie des énumérations* basée sur des propriétés rhétoriques et architecturales de ces structures (IRIT).

D'un point de vue informatique, les résultats obtenus sont :

1. *Un moissonneur de pages Web* (LIMSI). Cet outil permet de collecter des pages Web contenant des énumérations sur les moteurs de recherche Alta Vista et NorthernLight. Il interroge les moteurs de recherche au moyen de requêtes composées d'une sous-chaîne d'une amorce bien formée. Réalisé en Perl avec *wget*.
2. *Un analyseur de titres et énumérations* (LIMSI). Les énumérations contenues dans les pages Web sont décomposées en une amorce et une liste d'items. L'analyse s'appuie sur une grammaire des énumérations combinant de balises HTML et des marqueurs lexicaux signalant une énumération (*following, such as, list of...*). Réalisé en Perl.

3. *Un analyseur des initialisateurs (ou amorce)* (LIMSI). Les initialisateurs précédant les listes extraites des pages Web au moyen de l'outil précédent sont finement analysées au moyen de transducteurs en cascade. Réalisé en Perl.
4. Un outil de *représentation et manipulation d'images de texte* principalement dédié aux structures énumératives (IRIT). Réalisé en Java
5. des *patrons de fouilles* pour le recueil automatique des énumérations (ERSS). Réalisé avec le logiciel SATO.

La modélisation *linguistique* des structures énumératives et le modèle de l'architecture textuelle et de la Mise en Forme Matérielle (MFM) ont servi de fondement aux études *psycholinguistiques* (LPEQ, LORDAT) qui ont été, elles-mêmes, confrontées aux modélisations *informatiques*. D'un point de vue psycholinguistique, les expériences menées au cours de ce

projet ont fait avancer la connaissance des *processus cognitifs liés au traitement des énumérations selon leurs représentations spatiales*. Il apparaît, dans les résultats des laboratoires LORDAT et LPEQ, que les *variations typodispositionnelles ont une incidence sur la lecture et la mémorisation des énumérations*. Ces données portent sur des *adultes* (LPEQ) et sur des *enfants de 9 à 11 ans* (LORDAT). Ces premiers résultats encourageants incitent les différentes équipes impliquées dans ces expérimentations (IRIT, LORDAT et LPEQ) à poursuivre leur collaboration pour des recherches plus précises. Trois expériences sont prévues visant à confirmer par des études statistiques, les premiers résultats obtenus sur les enfants, et à mettre en évidence une stratégie de lecture particulière des adultes, lorsque la représentation spatiale d'une énumération est en adéquation avec sa structure sémantique.

Publications issues du projet

- Bush C. (2000), Analyse des déclencheurs des énumérations d'entités nommées sur le Web. Technical Report 2000-05, LIMSI-CNRS, mars.
- Bush C. (2002), Analysis of Named Entity Enumerations on the Web. Computer and the Humanities, Soumis et révisé en décembre 2001.
- Carrio L. (2001), *Incidence des variations typodispositionnelles sur la compréhension/mémorisation de textes de vulgarisation scientifique et le traitement cognitif des énumérations contenues dans ces textes par des enfants de 9 à 11 ans : élaboration d'un protocole expérimental*. Mémoire de DEA, Sciences du Langage, Université Toulouse-Le Mirail.
- Carrio L. (2002), : « Incidence des variations typodispositionnelles sur la compréhension/mémorisation de textes de vulgarisation scientifique et le traitement cognitif contenues dans ces textes par des enfants de 9 à 11 ans », Journées de PRESCOT, Inscriptio Spatiale du Langage: structures et processus, 29-30 Janvier 2002, IRIT Toulouse, 207-220.
- Dominguez C. (2001), *Structures spatio-linguistiques du texte : élaboration de textes expérimentaux pour un protocole de psycholinguistique*. Mémoire de DEA, Sciences du Langage, Université Toulouse-Le Mirail.
- Dominguez C. (2002), : « Étude de l'influence de la mise en forme matérielle d'un texte et d'un objet textuel particulier, la structure énumérative, sur leur traitement cognitif (compréhension/mémorisation) », Journées de PRESCOT, Inscriptio Spatiale du Langage: structures et processus, 29-30 Janvier 2002, IRIT Toulouse, 223-232.
- Garcia-Debanc C., Grandaty M (2001), Incidence des variations de la mise en forme textuelle sur la compréhension et la mémorisation de textes procéduraux (règles de jeux) par des enfants de 8 à 12 ans, *Langages* 141, mars 2001, 92 - 104.
- Garcia-Debanc C. (2001), « Les genres du discours procédural. Invariants et variations », *Pratiques* 110-111, 65-76.
- Garcia-Debanc C. (2002), : « Du modèle d'architecture textuelle à la déclinaison des textes à expérimenter auprès d'enfants de 9 à 11 ans : éléments méthodologiques et premiers résultats », Journées de PRESCOT, Inscriptio Spatiale du Langage: structures et processus, 29-30 Janvier 2002, IRIT Toulouse, 195-204.
- Grandaty M., Degeilh S., Garcia-Debanc C. (1998), « Rôle de la mise en forme matérielle dans la constitution du sens des textes à consignes : analyse psycho-linguistique et premiers résultats expérimentaux », in Virbel Jacques, Cellier Jean-Marie, Nespoulous Jean-Luc 1997 : *Cognition, Discours procédural, action*, PRESCOT, novembre 1997, 55 - 74.
- Grandaty M., Garcia-Debanc C. (2000), Les effets de la mise en page sur la compréhension et la mémorisation de règles de jeux par des enfants de 9 à 12 ans : conditions et résultats d'une étude expérimentale sur le texte, *Revue Paroles*, Mons, Langage et cognition.
- Jacquemin C. (2000a), L'Internet, un réservoir de langue infini? *EC-Arts*, 2
- Jacquemin C. et Bush C. (2000a), Fouille du Web pour la collecte d'Entités Nommées. In E. Wehrli, éditeur, Actes, TALN'2000, Lausanne, Suisse.
- Luc C., Mojahid, M., Virbel, J., Garcia-Debanc, C. & Péry-Woodley, M.-P. (1999), A linguistic approach to some parameters of layout: A study of enumerations. In R. Power & D. Scott, Eds., *Actes, AAAI 1999 Fall Symposium: « Using Layout for the Generation, Understanding or Retrieval of Documents »*, pp. 20-29. North Falmouth, Massachusetts.
- Luc C. (2000), Représentation et composition des structures rhétoriques et visuelles du texte. Approche pour la génération de textes formatés. *Thèse de doctorat*, Université Paul Sabatier. Novembre 2000.
- Luc C., Mojahid M., Virbel J (2001), *Système notational de l'architecture textuelle par image de page*. Dans CIDE2001 (4e Colloque International sur le Document Électronique), Toulouse, octobre 2001.
- Luc C., Virbel J. (2001), Le modèle de l'architecture textuelle: fondements et expérimentation. Dans *Verbum 23-1 « cohérence et relations de discours à l'écrit »*, 2001.
- Mojahid M., Jacques M.P., Sarda L (2001), *Repérer les structures du texte. Éléments pour la construction d'un modèle d'analyse*. Dans CIDE2001 (4e Colloque International sur le Document Électronique), Toulouse, octobre 2001.
- Pascual, E. & Péry-Woodley, M.-P. (1999), « Définition et action dans les textes procéduraux ». In J. Virbel, J.-M. Cellier & J.-L. Nespoulous (Eds.), *Cognition, Discours Procédural, Action*, Vol. II. Toulouse: PRESCOT, pp. 80-108
- Péry-Woodley, M.-P. (2001), « Modes d'organisation et de signalisation dans des textes procéduraux », *Langages*, 141, pp. 28-46.
- Schmid, S. (2001), *Représentation Organisationnelle et Compréhension: Rôle de la mise en forme matérielle dans la lecture*. Thèse de doctorat de l'université de Nice Sophia Antipolis.

Schmid, S. & Baccino, T. (1998), *Influence de la mise en forme des textes sur la compréhension*. Document Électronique, pp. 185-192 Europia Productions, Paris.

Schmid, S. & Baccino, T. (2001), *Stratégies de lecture pour les textes à consigne*, *Langages*, 141, 105-124.