



HAL
open science

Quelques aspects méthodologiques d'une étude de la fonction discursive des titres en corpus

Marie-Paule Jacques, Josette Rebeyrolle, Lydia-Mai Ho-Dac

► To cite this version:

Marie-Paule Jacques, Josette Rebeyrolle, Lydia-Mai Ho-Dac. Quelques aspects méthodologiques d'une étude de la fonction discursive des titres en corpus. Workshop ATALA "Modelling and describing discourse organisation in the age of the digital document", Jun 2004, , France. sic_00001227

HAL Id: sic_00001227

https://archivesic.ccsd.cnrs.fr/sic_00001227

Submitted on 7 Dec 2004

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.

Quelques aspects méthodologiques d'une étude de la fonction discursive des titres en corpus

Marie-Paule Jacques, Josette Rebeyrolle, Lydia-Mai Ho-Dac

Equipe de Recherche en Syntaxe et Sémantique, Université Toulouse-Le Mirail
Maison de la Recherche, 5, allées Antonio Machado, F- 31058 Toulouse
{mpjacques, rebeyrol, hodac}@univ-tlse2.fr

Nous voulons proposer ici une réflexion d'ordre méthodologique en rapport avec la construction d'un objet d'étude qui mêle typiquement étude du fonctionnement du discours et linguistique de corpus. Ce terrain sur lequel nous nous aventurons est celui de la compréhension de la ou des fonction(s) des titres au sein des documents écrits. Il s'agit de cerner la contribution des titres de sections et sous-sections à l'élaboration, l'organisation et la structuration du discours.

Les quelques études qui prennent les titres pour objet se centrent, par exemple, sur les titres d'œuvres littéraires (Hoek, 1981), de tableaux (Bosredon, 1997), ou les titres de presse (Sullet-Nylander, 1998) mais n'examinent pas les titres insérés dans les textes de documents.

Ceux-ci ont fait l'objet de premiers travaux (Hodac et al., 2004 ; Jacques, 2003 ; Rebeyrolle, 2003) à partir des hypothèses générales suivantes :

- il existe une variation de la forme et de la fonction des titres selon le genre de discours;
- les titres jouent un rôle dans la structuration du discours.

Pour tester ces hypothèses, nous nous sommes trouvées confrontées à la nécessité de définir une méthodologie et de déterminer quels éléments concrets observer dans les textes. Nous avons dû définir les « observables pertinents » à considérer pour élaborer – tout au moins commencer à élaborer – une description de la fonction des titres. Notre contribution se focalisera sur l'explicitation des choix et des décisions qui nous ont conduit à un certain nombre de descripteurs. Cette réflexion nous paraît être de nature à éclairer diverses questions méthodologiques qui se posent lorsque l'on souhaite appliquer les outils de la linguistique de corpus à une problématique de discours.

1. Un choix méthodologique essentiel : une étude de corpus

La description visée s'inscrit dans une linguistique de corpus¹, c'est-à-dire s'appuie, nous allons voir de quelle manière, sur l'analyse d'un ensemble de titres attestés, dans leur environnement textuel, dans des textes « écologiques » (i.e. ayant une réelle fonction sociale et non fabriqués pour les besoins de l'étude). L'inscription dans une linguistique de corpus tient, en partie, à l'hypothèse que les titres au sein des documents assurent un rôle cohésif, un rôle de liant textuel – entre autres fonctions, notamment la segmentation du texte. Pour déterminer de façon plus précise la validité d'une telle hypothèse, nous avons choisi d'analyser les titres dans des textes authentiques et non de fabriquer, sur la base de notre compétence de locuteur, des « textoïdes » plus aisés à manipuler.

Soulignons que, si l'objet d'étude implique un tel recours au corpus, celui-ci ne s'impose pas nécessairement. L. Hoek, qui vise l'élaboration d'une grammaire du titre d'œuvre littéraire, récuse fermement la méthode de corpus : « notre but n'est pas de faire des investigations sur un ensemble strictement délimité mais plutôt de décrire les invariants du titre [...] la délimitation d'un corpus décide d'avance des résultats : on n'y trouvera pas autre chose que ce qu'on y a mis. Une grammaire ne saurait être déduite à partir d'un corpus » (Hoek, 1981 : 19). C'est là le classique reproche adressé à une linguistique de corpus : elle ne décrit pas tous les faits possibles – ce serait pourquoi elle ne saurait autoriser la construction d'une grammaire – mais uniquement ceux qui sont présents dans le corpus et qui constituent, de façon inévitable, un « ensemble délimité ». De ce fait, la première décision à prendre tient donc au caractère inévitablement restreint d'un corpus et à la difficulté d'évaluer sa représentativité (Habert, 2000).

Pour limiter le risque de décrire des titres qui seraient spécifiques à un discours particulier, relevant d'un domaine et d'un registre, la solution adoptée consiste à procéder de façon contrastive sur trois

¹ Nous dirons même que notre étude se veut « corpus-driven » et non « corpus-based » (Tognini-Bonelli, 2001)

ensembles de textes² provenant de domaines et de genres différents : des articles scientifiques de deux domaines distincts, l'ingénierie de la connaissance et la géopolitique ; des documents de travail de professionnels de la gestion des déplacements. Ces textes ont été choisis en raison du fort rôle d'organisation et de structuration du discours qu'y jouent les titres, tout en ne relevant pas pour autant tous trois du même genre discursif. Nous faisons également l'hypothèse qu'en réalisant l'analyse sur plusieurs corpus et non un seul, se feront jour les éléments de stabilité et de variation des fonctions des titres. Ainsi, la multiplication et l'instauration d'une possibilité de comparaison des corpus nous paraissent favorables d'une part, à l'observation des caractéristiques spécifiques que peuvent présenter les titres dans tel ou tel ensemble de textes et, d'autre part, à une quantification contrastive des phénomènes analysés. Nous revenons plus loin sur la question de la quantification.

Dans cet objectif et pour ne pas avoir à extrapoler nos observations, nous avons choisi de ne pas travailler sur des corpus de tailles différentes mais plutôt de niveler la taille des corpus à l'intérieur des trois groupes. Pour ce faire, il est nécessaire de déterminer le critère le plus pertinent à prendre en compte pour la constitution de corpus comparables quantitativement : est-ce que la base de comparaison pertinente est le nombre de mots, le nombre de textes ou le nombre de titres de chaque corpus ? L'un ou l'autre conduit à une constitution de corpus différente ; en effet, – et ce sera un point à noter dans l'étude proprement dite – les textes dans chacun des trois domaines présentent un ratio nombre de mots/nombre de titres différent. Si on décide, par exemple, de constituer des corpus de taille équivalente en nombre de mots – ce qui est une pratique habituelle en linguistique de corpus, alors on se retrouve avec un nombre de titres qui varie du simple au double et on est obligé d'extrapoler. Il semble donc plus pertinent de privilégier le critère du nombre de titres pour avoir un nombre comparable d'observables dans chaque corpus. Ce choix se justifie également par le fait que c'est particulièrement le comportement du titre dans le texte qui nous intéresse.

2. Délimitation des observables pertinents

Une fois la question de la délimitation du corpus tranchée, se pose ensuite celle de la délimitation de ce qui va être analysé, autrement dit des observables pertinents. On est là devant la difficulté inhérente à une étude qui se donne pour objet l'analyse de faits discursifs : le discours n'est pas en soi un observable, car il met en jeu bien plus que ce qui est exprimé à la surface des textes, dans la mesure où il suppose de l'interprétation. Et qui dit interprétation dit construction d'un modèle mental (Fauconnier, 1984) relevant d'une interaction de processus tels que l'établissement et le maintien d'un thème (Fries, 1995), la mise en relation d'unités textuelles (Mann et Thompson, 1988), etc. qui impliquent la mobilisation de connaissances de diverses sortes, tout à la fois linguistiques et extra-linguistiques (cf. Cornish, 2003). Il est donc souvent délicat de constituer des observables, étant donné que les faits de discours ne s'organisent pas tous de façon univoque sur des marques formelles, facilement repérables. Or, si l'on veut mettre en œuvre des méthodes d'analyse de corpus, et plus particulièrement des traitements automatiques, il importe de se donner des éléments formels comme support des observations.

La reprise et l'annonce comme indices formels des fonctions du titre

Notre point de départ étant déjà un objet repérable sur des critères formels : les titres sont reconnaissables dans les textes par un certain nombre de propriétés formelles – mise en forme matérielle (gras, italique, majuscule, etc.), disposition... –, nous faisons l'hypothèse qu'il est possible de saisir leur contribution à l'organisation du discours et plus particulièrement d'évaluer leur rôle cohésif par la prise en compte d'éléments formels. Car la cohésion au sein d'un texte (cf. Halliday et Hasan, 1976) s'appuie, au moins en partie, sur un chaînage référentiel, lui-même réalisé par la répétition d'items lexicaux, la pronominalisation, etc.. Nous pourrions donc cerner la contribution des titres à la cohésion textuelle en prenant l'exacte mesure de la récurrence dans le texte des éléments qui les composent, ceci s'appréciant par un regard autant sur ce qui suit le titre, sa reprise éventuelle après son occurrence, que sur ce qui le précède, son annonce éventuelle dans le texte antérieur. Ces éléments ne sont pas si simples à prendre en considération car il s'agit de décider de ce qu'on va estimer être une annonce ou une reprise du titre. Par exemple, dans l'extrait suivant :

² Précisons que tous ces documents sont sous format numérique pour l'étude, ce qui autorise la mise en œuvre de traitements automatiques, mais que ce sont à l'origine des documents écrits pour être diffusés sous format papier. Dans notre étude, l'appellation « numérique » s'applique uniquement à la forme sous laquelle nous disposons de ces documents et ne concerne aucunement une modalité de communication.

L'unité de fonctionnement et de développement

L'unité de fonctionnement et de développement a pour rôle de préparer les scénarios de gestion du trafic et de s'assurer du bon fonctionnement du dispositif mis en place. Ses activités sont les suivantes :

L'activité Développement :

Elle consiste en l'élaboration des propositions de politique d'exploitation et de scénarios d'exploitation soumis, pour décision, au Comité d'Exploitation.

Deux analyses seraient ici possibles. On peut considérer que le second titre *L'activité Développement* est annoncé par la phrase qui le précède *Ses activités sont les suivantes*, parce qu'on y retrouve un même nom – *activité*. Mais on peut aussi estimer le contraire, parce que la phrase qui précède ne comporte pas strictement les mêmes éléments lexicaux. A quels indices donner priorité ? Ou encore, dans ce second extrait :

Composition de l'équipe ERATO

En résumé, l'équipe pourrait comprendre :

Le rapport sémantique indéniable entre le titre et la première phrase du paragraphe qui le suit doit-il être étiqueté « reprise » ou... quoi ? Il y a dans cette séquence une reprise anaphorique du terme complexe « l'équipe ERATO » par sa tête « l'équipe » (Jacques, 2003), mais que dire au niveau du titre dans son entier ?

On le voit, l'élaboration de critères d'analyse ne va pas de soi.

Difficultés d'appréhension de relations sémantico-pragmatiques

On peut également s'appuyer sur les travaux existants sur l'anaphore et les chaînes de référence, mais, pour notre propos qui est de cerner le rôle du titre dans l'organisation globale du texte, ils s'avèrent vite incomplets. En effet, si pour des raisons didactiques, les exemples que nous venons d'évoquer couvrent des empan de textes relativement étroits (co-texte immédiatement antérieur ou postérieur), ils ne montrent qu'une partie des relations qu'un titre peut établir avec le reste du texte. Si l'on restreint l'observation à l'environnement textuel immédiat du titre, on ne voit le rôle de celui-ci qu'en tant que participant à une chaîne de référence locale. Or, il ne semble pas du tout impensable de voir surgir à la page 15 d'un document un titre en rapport avec ce qui aura été mentionné à la page 3, ou de voir un titre de niveau inférieur être constitué d'une partie de titre de niveau supérieur, ou de voir s'établir d'autres relations entre titres. On n'est plus là dans le cadre de la contribution du titre à la construction de référence et de chaînes de référence dans le texte, mais bien dans le cadre de sa contribution à l'organisation globale du discours. Ce type de phénomène est assez difficile à étudier et même tout simplement à voir, car, en général, quand on se préoccupe de discours, on est soit dans une analyse proche du texte mais à un niveau très local (par exemple la relation entre segments de textes adjacents), soit à un niveau global mais dans une analyse *abstraite* du texte, à visée computationnelle (segmentation automatique ou repérage de zones selon des critères statistiques ou numériques, portant sur la distribution des mots et leur récurrence, cf. Hearst, 1997). Il nous importe de nous donner les moyens d'observer ces relations non seulement locales mais aussi distantes des titres avec le reste du texte, car nous faisons l'hypothèse que ce sont elles qui font la spécificité de la structuration opérée par les titres.

Quantification des observables

Le rôle des titres dans le fonctionnement du discours s'apprécie non seulement en examinant les modalités de son inscription dans le texte, mais aussi en quantifiant les différents facteurs que nous venons d'évoquer. C'est là que l'importance d'avoir un nombre équivalent de titres dans chaque corpus apparaît. Par exemple, s'il se dégage de l'analyse que dans l'un des corpus, les titres de niveaux les plus bas sont significativement repris après leur occurrence, cela permet de cerner de façon plus fine et plus immédiate la contribution des titres. Cela peut éventuellement aussi permettre d'évaluer une possible influence du domaine et du genre sur le rôle des titres.

Globalement, une quantification précise des diverses observations peut permettre de faire émerger des corrélations ou non des différents facteurs analysés.

3. A titre de bilan provisoire...

L'une des difficultés de notre approche, et c'est en cela qu'elle nous semble pouvoir offrir matière à réflexion et à discussion, est qu'elle ne trouve pas à s'appuyer sur un cadre théorique adéquat. Les différents modèles du discours sont en effet en partie inappropriés dans le sens où ils sont conçus pour rendre compte des propositions et des relations entre propositions ; de bons exemples sont la Théorie du Centrage (Walker, Joshi & Prince, 1998) ou encore la Théorie des Structures Rhétoriques (Mann & Thompson, 1988). Ils n'offrent pas un cadre satisfaisant pour analyser la structuration de textes de plusieurs pages tels que ceux qui constituent notre corpus. De plus, le discours étant un lieu d'interaction systématique de phénomènes, on est souvent obligé de faire appel à telles ou telles élaborations théoriques en ordre dispersé, ce qui peut parfois apparaître comme un « bricolage » peu rigoureux.

De son côté, la linguistique de corpus considère encore souvent les textes comme des « paquets de mots », au mieux des « paquets de phrases », et est bien loin d'offrir une modélisation de faits discursifs. En effet, encore peu de travaux prennent en compte la localisation dans le texte des marques formelles, à l'exception de quelques-uns comme par exemple ceux de Virtanen (1992), qui accorde un statut d'organisateur de niveau global aux adverbiaux placés en initiale de paragraphe.

Il semble qu'à l'heure actuelle, les difficultés que nous évoquons soient le lot commun de nombre d'études qui prennent pour objet des phénomènes discursifs, car l'analyse en discours ne correspond pas à l'extrapolation de l'analyse en phrase, mais suppose un changement de point de vue : l'observation du texte tant dans sa globalité que dans sa linéarité, sa localité.

Références

- Bosredon, B. (1997). *Les titres de tableaux. Une pragmatique de l'identification*. Paris: P.U.F.
- Cornish, F. (2003). The roles of (written) text and anaphor-type distribution in the construction of discourse. *Text* 23 (1), pp.1-26.
- Fauconnier, G. (1984). *Espaces mentaux*. Paris: Les éditions de Minuit.
- Fries, P. (1995). Themes Method of Development and texts. R. Hasan & P. Fries (eds) *On Subject and Theme: A Discourse Functional Perspective*. Amsterdam/Philadelphia: John Benjamins, pp.317-359.
- Habert, B. (2000). Des corpus représentatifs : de quoi, pour quoi, comment ? *Cahiers de l'Université de Perpignan*, 31, pp.11-58.
- Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. Londres: Longman.
- Hearst, M.A. (1997). TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages. *Computational Linguistics*, 23 (1), pp.33-64.
- Ho-Dac, M., Jacques, M.-P. & Rebeyrolle, J. (2004). Sur la fonction discursive des titres, Regards croisés sur l'unité texte, Nicosie, Chypre.
- Hoek, L.H. (1981). *La marque du titre*. La Haye: Mouton.
- Jacques, M.-P. (2003). *Approche en discours de la réduction des termes complexes dans les textes spécialisés*. Doctorat Nouveau Régime, Université Toulouse-Le-Mirail.
- Mann, W.C. & Thompson, S.A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3), pp.243-281.
- Rebeyrolle, J. (2003). Forme linguistique et fonction discursive des titres de section. Colloque AFLS, Tours.
- Sullet-Nylander, F. (1998). Le titre de presse. Analyses syntaxique, pragmatique et rhétorique. Stockholm, Stockholms universitet.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam / Philadelphia: John Benjamins.
- Virtanen, T. (1992). *Discourse Functions of Adverbial Placement in English. Clause-Initial Adverbials of Time and Place in Narratives and Procedural Place Descriptions*. Abo Akademi: University Press.
- Walker, M.A., Joshi, A.K. & Prince, E.F. (1998). *Centering Theory in Discourse*. Oxford: Clarendon Press.