



HAL
open science

Les séries linéaires dans le discours

Agata Jackiewicz

► **To cite this version:**

Agata Jackiewicz. Les séries linéaires dans le discours. Langue française, 2005, 148 (148), pp.95-110.
10.3917/lf.148.0095 . hal-01839749

HAL Id: hal-01839749

<https://hal.science/hal-01839749>

Submitted on 16 Jul 2018

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

LES SÉRIES LINÉAIRES DANS LE DISCOURS

Agata Jackiewicz

Armand Colin | « Langue française »

2005/4 n° 148 | pages 95 à 110

ISSN 0023-8368

ISBN 9782035770868

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-langue-francaise-2005-4-page-95.htm>

Pour citer cet article :

Agata Jackiewicz, « Les séries linéaires dans le discours », *Langue française* 2005/4 (n° 148), p. 95-110.

DOI 10.3917/lf.148.0095

Distribution électronique Cairn.info pour Armand Colin.

© Armand Colin. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Agata Jackiewicz

Laboratoire LaLICC

UMR 8139 (CNRS - Université Paris-Sorbonne)

Université de Paris-Sorbonne (ISHA)

28, rue Serpente 75006 Paris – France

Agata.Jackiewicz@paris4.sorbonne.fr

Les séries linéaires dans le discours

I. INTRODUCTION

Tout discours résulte d'un processus d'énonciation qui découpe et articule des contenus. Ces contenus peuvent être liés, divisés, enchâssés, mis en valeur suivant les choix et objectifs de l'auteur. Un énonciateur coopératif peut augmenter très nettement la lisibilité de son discours en rendant saillantes certaines de ses caractéristiques formelles ou fonctionnelles. Il dispose à cet effet de moyens typo-dispositionnels et discursifs (découpage en paragraphes, diverses « balises », commentaires autonymes, mots de liaison...), lesquels donnent au co-énonciateur, dans une visée résolument dialogique, des instructions pour la gestion de l'activité de lecture et de compréhension.

Lors de plusieurs travaux en ingénierie linguistique sur le résumé et la fouille sémantique de textes (Minel *et al.*, 2001 ; Crispino *et al.*, 2004), nous avons rencontré la problématique posée par le découpage de documents en unités de traitement et l'identification de la configuration de ces unités. Annoter sémantiquement des objets textuels (une définition, une citation, une séquence causale...), pour (i) parcourir le texte en suivant ces objets (navigation intradocumentaire), ou pour (ii) en extraire certains, conformément à un critère sémantique ou à un jugement d'importance (synthèse sélective), nécessite au préalable une étude de la manière dont les frontières de ces objets et leurs modes d'organisation (succession, enchâssement...) sont signalés dans et par le discours.

La *sériation*, qui est un procédé fondamental d'articulation textuelle permettant à la fois d'ordonner et de situer sur un axe d'équivalence paradigmatique des portions de texte, s'est imposée rapidement comme l'objet de traitements transversaux à plusieurs applications informatiques. De plus, l'intérêt théorique de cette opération sur le plan de l'analyse du discours est évident. La sériation dans le discours est un procédé énonciatif par lequel l'énonciateur informe le

co-énonciateur qu'il donne à son propos la forme d'une suite ordonnée d'unités équivalentes par rapport à un critère, ce qui peut être paraphrasé par la formule : « je dis que ce qui est dit forme une succession ordonnée d'éléments de même nature ». Ainsi dans la séquence (1), les adverbiaux *Primo* et *Deuzio* fonctionnent de concert pour introduire respectivement le premier et le deuxième item, au sein d'une série de deux éléments. La longueur de la série et le principe fédérateur qui unit ses composants sont explicitement annoncés dans l'énoncé « amorce » : *Deux éléments fondent...*

- (1) *Deux éléments fondent le plan de financement choisi par le gouvernement pour réussir sa réforme des retraites. Primo, a encore rappelé François Fillon, ministre des affaires sociales, devant les parlementaires de l'UMP le 5 mai, « il n'existe pas de trésor caché » qui permettrait de raser gratis. Deuzio, il est hors de question de revenir sur la promesse de Jacques Chirac de réduire, ou tout au moins de stabiliser, le niveau des prélèvements obligatoires.* (13 mai 2003, p. 2, *Le Monde*)¹

Force est de constater que les traces (tant physiques que linguistiques) de cette organisation « dialogique » du discours sont encore peu étudiées en linguistique théorique. L'opération complexe de sériation constitue également un puissant facteur de cohésion textuelle, combinant les relations de connexion entre items de la série et les relations d'indexation, créatrices d'items en tant que blocs homogènes et distincts. Dans cette optique, il est intéressant d'aborder l'étude des marques impliquées dans l'expression des séries du point de vue de leur capacité simultanée à connecter des segments textuels et à les indexer (ou cadrer).

Dans la présente contribution, nous nous proposons de montrer à partir d'une expérience complète, allant d'une analyse linguistique à l'implantation, menée sur l'objet textuel « série linéaire » (Jackiewicz, 2002 ; Jackiewicz et Minel, 2003 ; Couto *et al.*, 2004), comment l'objectif du repérage automatique d'objets et de structures de discours (i) oriente le travail linguistique de collecte et d'analyse de marques d'organisation textuelle, (ii) questionne le cadre méthodologique, faisant notamment apparaître la nécessité de représenter formellement et de traiter informatiquement des faisceaux d'indices contextuels, relatifs à plusieurs opérations discursives en présence, (iii) amène à considérer la question des normes et des usages à l'œuvre dans les corpus textuels exploités.

Nous focalisons notre attention sur les phases de recueil et d'analyse des marques et des connaissances linguistiques impliquées dans l'expression des séries, afin de montrer que ce travail nécessite conjointement (i) le recours à des corpus textuels, car les manuels de grammaire ne disent pas comment construire et signaler une série ; (ii) des analyses linguistiques « classiques », car les exemples attestés ne sont pas suffisants, aussi est-il nécessaire de procéder à des manipulations paradigmatiques pour faire apparaître des contraintes et des régularités d'expression ; (iii) la prise en considération des propriétés de

1. Dans les extraits textuels qui illustrent nos propos, nous mettons en gras les introducteurs d'item et soulignons les éléments pertinents des amorces et des rétro-évaluations.

nature logique et cognitive de la sériation, ce qui aide à cerner la nature de cette opération dans le discours et, par conséquent, à dominer la diversité des marques linguistiques impliquées dans la sériation discursive.

L'ensemble du travail s'inscrit dans la lignée des recherches sur l'encadrement du discours (Charolles, 1997) et emprunte le cadre formel de la méthode d'exploration contextuelle (Desclés, 1997). Nous avons exploité plus ponctuellement des travaux portant sur les marques d'intégration linéaire et les énumérations (Turco et Coltier, 1988 ; Péry-Woodley, 2000 ; Luc et Virbel, 2001), ainsi que sur le métalangage (Borillo, 1985 ; Authier-Revuz, 1995 ; Rey-Debove, 1997 ; Desclés et Guentchéva, 2000).

2. LES SÉRIES LINÉAIRES : OBSERVATION EN CORPUS

Comment sont construites et de quelle manière sont signalées les séries linéaires dans le discours ? Nous avons cherché à répondre à cette question en analysant des séries textuelles extraites des documents journalistiques (issus du *Monde diplomatique* et du *Monde*²) et en consultant ponctuellement la base de données textuelles Frantext. Nous avons collecté 311 séries différentes³, pour en analyser les caractéristiques morphologiques d'une part, et les moyens d'expression, de l'autre. Pour le recueil des séries, nous sommes partie d'une acception intuitive de cet objet, selon laquelle une série est composée d'au moins deux items. Une série est identifiée grâce aux introducteurs de ses items, dont au moins un doit être une marque d'intégration linéaire « classique » telle que {*d'une part, d'autre part, de l'autre ; d'un côté, d'un autre côté ; en premier lieu, en second lieu, ... en dernier lieu ; d'abord, ensuite, enfin ; dans un premier temps, dans un second temps... ; premièrement, deuxièmement... ; primo, secundo, tertio...*}.

2.1. Une série « typique » et ses variantes

Il ressort de cette analyse que les séries textuelles balisées par des marques discursives sont (i) typiquement constituées de 2 ou de 3 items (avec une longueur maximale attestée de 10 items, mais théoriquement illimitée pour certaines catégories d'introducteurs), dont la portée peut aller d'un syntagme à une suite de paragraphes ; (ii) structurées à un niveau, plus rarement à deux niveaux (emboîtement) ; (iii) précédées par une amorce explicite ou reconstruite *a posteriori* ; (iv) parfois clôturées par une rétro-évaluation portant sur l'ensemble de la série.

2. Le corpus issu du *Monde diplomatique* (LMD) est constitué par 12 520 documents parus entre 1984 et 1998 et disponibles sur un cédé-rom. Celui qui est extrait du *Monde* représente l'édition complète des années 2002 et 2003 sur cédé-rom. Ce corpus a été choisi en raison du faible degré de structuration visuelle (dispositionnelle et typographique) des textes qui le composent.

3. Ce qui représente 311 suites d'introducteurs discursifs différentes {(1) *d'une part, de l'autre ; (2) en premier lieu, en second ; ... ; (311) primo, deusio*}

Sur le plan de la signalisation discursive, (i) les items sont généralement introduits par des marqueurs d'intégration linéaire (MIL) (*en premier lieu, premièrement, d'abord...*), dont les configurations composent assez librement des introducteurs issus de la numération et des paradigmes temporels et spatiaux⁴, mais qui peuvent également être incomplètes ou partiellement relayées par d'autres marques (*mais, parce que...*) ; (ii) l'amorce explicite contient généralement un classificateur (*élément, étape, composante...*) et un quantificateur (*deux, double, triple...*), qui assurent d'une manière complémentaire aux MIL le signalement des items (leur paradigme commun, leur nombre) ; (iii) la rétro-évaluation, lorsqu'elle est présente, est introduite par des marques de reformulation (*en bref, autrement dit, en définitive...*) ou de quantification totalisante (*tous, au total, les deux, l'ensemble...*).

La nature proprement discursive des introducteurs exerce une incidence sur les caractéristiques morphologiques de la série. Généralement plus courtes et moins complexes dans leur structure que leurs homologues « visuelles »⁵, les séries balisées par les marqueurs d'intégration linéaire sont aussi plus difficiles à identifier, en raison de la variabilité et de la polysémie des formes qui peuvent jouer le rôle des organisateurs discursifs. On constate toutefois que plus une série est longue et/ou plus ses items sont de taille importante, plus le signalement discursif est saillant (y compris visuellement) et régulier.

Ce résumé des propriétés caractéristiques des séries linéaires dans le discours, qui donne l'impression d'une variabilité difficile à dominer, réfère à une somme de données soumises à l'influence d'au moins trois facteurs. Premièrement, la morphologie d'une série est susceptible de variations. Deuxièmement, pour une série donnée, il existe plusieurs possibilités logiquement équivalentes entre elles, mais sémantiquement et stylistiquement différentes, permettant de la signaler. Enfin, sur ces deux plans (de la morphologie et de l'expression), le respect ou non de certains critères de lisibilité (qu'ils soient « officiels » et imposés par des normes éditoriales, ou subjectifs dictés par des choix personnels de l'auteur), a une incidence sur l'effort de lecture nécessaire pour délimiter et interpréter correctement une série.

2.2. Signalement des séries : règles prescriptives et faits statistiques

Le genre textuel de notre corpus exerce-t-il un impact sur la morphologie et le signalement des séries qui y sont exprimées ? Les textes de communication, nous le savons, sont soumis à des règles d'écriture et de mise en forme destinées à faciliter leur réception. Le guide rédactionnel du *Monde* (*Le style du*

4. Il s'agit des séries « régulières », telles que *d'une part, d'autre part ; d'un côté, d'un autre côté ; en premier lieu, en second lieu, ... en dernier lieu ; d'abord, ensuite, enfin ; dans un premier temps, dans un second temps... ; premièrement, deuxièmement... ; premier, deuxième... ; primo, secundo...*

5. Nous faisons référence ici aux séries introduites par des balises typographiques (tiret, puce, numéro).

Monde, 2002) expose dans le chapitre « Le contrat de lecture » plusieurs règles visant à augmenter la clarté, la pédagogie et la précision des articles⁶. Cependant, ce guide ne fournit pas de recommandations sur la structuration des textes au niveau global, bien qu'indirectement, le principe de la pyramide inversée⁷ qui s'applique à certaines catégories d'articles, exerce une incidence sur l'organisation des contenus. Il existe toutefois des règles, non consignées dans ce document, que certains journalistes⁸ affirment avoir apprises lors de leur formation et appliquer systématiquement dans leurs écrits. Citons en particulier la règle stipulant que toute énumération doit être introduite d'une manière régulière (« si j'écris *d'une part*, je dois poursuivre par *d'autre part* ou par *de l'autre* ») ou encore celle qui régit l'emploi de l'adjectif *second*, en le réservant aux énumérations composées de deux éléments. Nous avons vérifié l'application de ces deux règles sur un échantillon de 200 séries tirées du *Monde*. Notre conclusion est que ces règles d'écriture, non seulement ne sont pas communément respectées, mais pour chacune d'entre elles, c'est l'inverse qui tend à constituer la norme. En effet, concernant la régularité d'expression, seulement 18 % des séries possèdent des balises régulières. Quant à la deuxième règle, on constate que dans 40 séries sur 160 séries de longueur supérieure ou égale à trois items, l'introducteur du deuxième item comporte l'adjectif *second*.

Le tableau 1 présente quelques exemples de séries d'introducteurs attestées dans le corpus du *Monde*. Toutes hétérogènes, certaines d'entre elles étant incomplètes (le symbole 0 représente une balise manquante), les séries citées ne peuvent de toute évidence être qualifiées de régulières. L'emploi de l'adjectif *second* dans une série de 4 éléments ou plus y est exemplifié trois fois. Par ailleurs, on note parmi ces introducteurs la présence des formes qualifiées par (Grevisse, 1993) de « barbares » (*deusio, deuzio, deuxio...*). Ajoutons que ces formes figurent tant dans des articles rédigés par des journalistes « maison », que dans des écrits signés par des correspondants ou des experts extérieurs.

6. Par exemple, pour transmettre clairement les informations, il est entre autres conseillé aux journalistes de (i) préférer la voie active à la voie passive, (ii) adopter la forme affirmative de préférence, (iii) rédiger des phrases courtes en évitant des rafales de subordonnées, (iv) proscrire les incises à répétition...

7. « La structure de la pyramide inversée. Une dépêche d'agence commence par l'essentiel et va ensuite au moins important, en paragraphes successifs. Ceux-ci doivent pouvoir être retranchés, en commençant par la fin de la dépêche, sans que le fond même de l'information soit dénaturé. » (*L'agencier*, Le manuel AFP, Agence France-Presse, 2004, p. 17).

8. Il s'agit d'un journaliste du *Monde* que nous avons interrogé personnellement.

| 1 | 2 | 3 | 4 | 5 | 6 |
|------------------------|-------------------|----------------------|--------------------------|----------------|--------|
| En premier lieu, | En second, | En troisième lieu, | En quatrième lieu, | Ensuite, | Enfin, |
| 0 | 0 | Troisièmement | Quatrièmement | Enfin | |
| ... d'abord | Ensuite | Troisièmement, | Quatrièmement, | Cinquièmement, | Enfin, |
| D'abord, | En second lieu, | Troisièmement, | Enfin, | | |
| La première | La seconde | Troisièmement | Enfin, | | |
| Premièrement, | Deuxio, | Troisièmement, | | | |
| 0 | Ensuite, | Troisièmement, | Enfin, | | |
| Dans un premier temps | Deuxièmement | Troisième source | | | |
| Primo, | 0 | , en troisième lieu, | | | |
| Primo | En deuxième lieu, | Troisième grief | | | |
| Primo | Secundo | | | | |
| Primo | Secundo | Tercio, | | | |
| Primo | Deuzio | Tertio, | | | |
| Primo | Deusio | | | | |
| 0 | Secundo, | Tertio, | Quarto, | Enfin, | |
| Premier commandement : | Deuxio, | Tertio, | | | |
| 0 | 0 | Tertio, | La quatrième conséquence | | |

Tableau 1 : Séries d'introducteurs attestées dans le corpus du *Monde* (extrait)⁹.

3. LES SÉRIES LINÉAIRES À LA LUMIÈRE DES ANALYSES LINGUISTIQUES

Les données collectées sur corpus révèlent les procédés morphologiques et sémantiques de constitution de balises (introducteurs d'items). Elles donnent également des indications sur d'éventuelles contraintes qui peuvent régir l'association des balises au sein d'une série particulière. Sur cette base, il devient possible de prédire des séries non attestées dans le corpus, mais réalisables en pratique. Les séries attestées livrent par ailleurs un savoir précieux sur les conditions contextuelles dans lesquelles une forme *a priori* polysémique ou

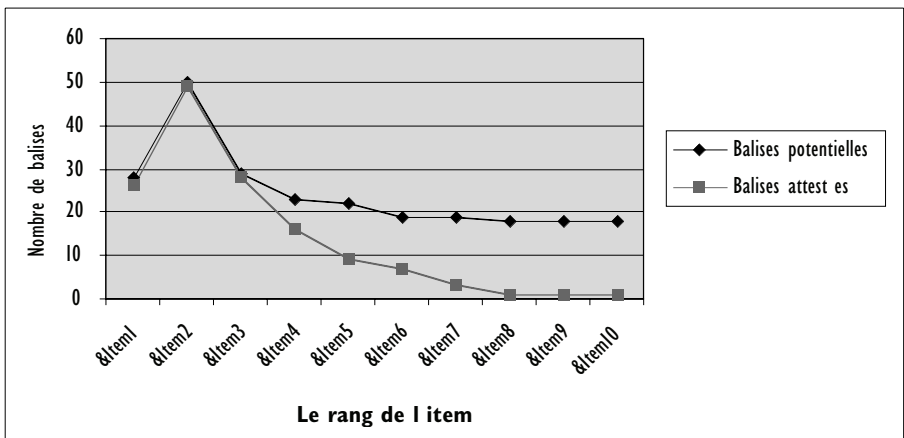
9. L'emploi des majuscules et la présence des signes de ponctuation dans le co-texte immédiat des introducteurs ayant été consignés dans la base des séries attestées, nous les restituons ici, sans qu'ils soient significatifs pour l'argumentation en cours.

polyfonctionnelle telle que *d'une part, premièrement* ou *enfin* pourra être interprétée comme un introducteur potentiel d'item au sein d'une série.

3.1. Création de paradigmes d'introducteurs

Les introducteurs attestés au sein des différentes séries prélevées dans notre corpus ne représentent pas toutes les balises possibles. Il est donc nécessaire d'enrichir les paradigmes amorcés, regroupant successivement tous les introducteurs possibles pour le premier, le deuxième... le n-ième item de la série. Chaque paradigme enferme les marques logiquement équivalentes, car elles réfèrent au même rang dans la série, mais différentes sur les plans morphologique et sémantique.

Plusieurs procédés permettent de compléter les paradigmes constitués. Tout d'abord, la plupart des séries régulières étant potentiellement infinies, on est amené à ajouter, aux formes attestées pour les k premiers items d'une série, les formes existantes qui vont au-delà du k-ième item. Ainsi, on « prolongera » la série {*primo, secundo, tertio*} par les termes {*quarto, quinto, sexto, septimo...*}. En deuxième lieu, il convient de systématiser pour certains introducteurs les formes sensibles au genre, ainsi que les déterminants qui peuvent leur être associés. De cette façon, si l'on atteste l'existence des syntagmes *ce premier/le premier/un premier*, on peut compléter le paradigme du rang 1 par les formes *cette première, la première/une première*. En troisième lieu, il est nécessaire de vérifier pour certains introducteurs composés la possibilité d'y insérer des éléments adverbiaux ou adjectivaux (*en premier lieu/en tout premier lieu...*) en surveillant les conséquences sur le plan sémantique que cela entraîne. Notons enfin que les quelques rares occurrences d'introducteurs verbaux attestées (*commencer par, se terminer par...*) soulèvent la question de savoir s'il est opportun d'accepter comme balises potentielles leurs synonymes les plus courants.



Graphique I : Introducteurs attestés en corpus versus introducteurs possibles présentés par rang de l'item.

À l'issue de ce travail, il devient possible de ré-examiner les ressources attestées à la lumière de celles qui sont attestables (voir le graphique 1). Nous constatons qu'il existe en moyenne 20 possibilités différentes pour introduire un item. Ce nombre est sensiblement plus important pour le deuxième item (49 formes différentes, toutes attestées), ce qui traduit un besoin fréquent d'exprimer une opposition ou plus simplement une nette différenciation entre deux unités. Les séquences formées de deux éléments se caractérisent donc sur le plan stylistique par une grande diversité d'introducteurs possibles, tous effectivement employés. Le deuxième fait remarquable est que dans le cas des séries longues (quatre items ou plus), peu de balises disponibles sont effectivement exploitées par les auteurs. Les séries longues tendent à être régulières (certainement pour mieux soutenir la mémoire du lecteur), ce qui se manifeste entre autres par l'emploi préférentiel des introducteurs « réguliers ». Deux « familles » d'introducteurs sont utilisées de préférence pour introduire de longues énumérations d'items discursifs : {*premièrement, deuxièmement, troisièmement...*} et {*le premier, le deuxième, le troisième...*} avec, pour cette dernière, plusieurs variantes possibles liées au genre et à la détermination.

3.2. Typage des introducteurs

Une fois les paradigmes d'introducteurs constitués, il est possible de les décrire sur les plans morpho-syntaxique, sémantique et fonctionnel. Plusieurs critères, intervenant conjointement ou de manière indépendante, permettent de distinguer :

- les balises d'ouverture (*d'abord, en premier lieu...*), de relais (*ensuite, en second lieu...*) et de clôture (*en dernier lieu, enfin...*) ; on note que pour signaler le dernier item de la série, les auteurs n'hésitent pas à composer des introducteurs « redondants » à partir de plusieurs indices de clôture (*enfin le dernier élément...*) ;
- les balises indépendantes (*en premier lieu, premièrement...*) et les balises liées (*le premier objectif, troisième composante...*), la liaison s'établissant entre l'item et l'amorce par l'intermédiaire d'un classificateur (*objectif, composante...*) ;
- les balises simples (*troisièmement, enfin...*) et les balises composées (*enfin le troisième élément*) ;
- les balises neutres (rôle uniquement organisationnel : les marques d'intégration linéaire) et les balises bi-fonctionnelles (rôle organisationnel et argumentatif : *mais, parce que...*) ;
- les balises discursives, les balises typographiques (tiret, puce, numéro), les balises typo-discursives (*- premièrement... ; - deuxièmement...*).

Ces distinctions sont importantes, car elles permettent d'adapter finement les heuristiques de repérage aux différentes catégories d'introducteurs. Par exemple, une balise bi-fonctionnelle est recherchée quand on ne trouve pas à l'intérieur d'un espace de recherche correspondant à un item donné de marque d'intégration linéaire susceptible de l'introduire. Comme on le voit dans (2), la

deuxième occurrence du connecteur *parce que* introduit (comme la première) un argument étayant une position, mais elle assure également une fonction de balise.

- (2) *Et pourtant, de plus en plus nombreux sont ceux qui dénoncent ce système comme une imposture. En premier lieu parce que, en Europe, il a toléré l'apparition de 20 millions de sans-emploi et de 50 millions de pauvres... Parce que certains États en viennent à accepter une sorte de tiers-mondisation de leurs sociétés. (LMD octobre 96, p. 1)*

3.3. Propriétés contextuelles des introducteurs

Les marques discursives qui introduisent les items des séries linéaires sont dans leur majorité polysémiques et polyfonctionnelles (*d'abord, d'une part, premier, d'un côté...*). En tant que candidat introducteur d'item, une occurrence de forme doit (i) vérifier les conditions (locales) dans lesquelles elle devient un introducteur de cadre de discours (cadre organisationnel, en l'occurrence), (ii) être mise en relation avec d'autres marques qui charpentent la structure en série (marques potentiellement introductrices d'items, équivalentes et complémentaires ; classificateurs au sein d'une amorce...) – nous explicitons les différents cas de figure possibles dans le § 5.

En qualité d'introducteur de cadre, le marqueur ne doit pas être syntaxiquement intégré à l'énoncé où il figure matériellement, ce qui contraint l'ensemble de ses positions possibles. En général, il figure en initiale détachée, souvent absolue, ou dans une incise. Un certain nombre de marques significatives peuvent se trouver dans son contexte proche (gauche, plus rarement droit), comme on le voit dans l'exemple (3). Dans une architecture en série, il s'agit le plus souvent (i) des signes de ponctuation tels que deux-points ou point-virgule, (ii) de certains connecteurs (*mais, et, surtout, parce que...*) ou marqueurs de clôture (*enfin, finalement...*), (iii) de divers marqueurs tels que *si, c'est, comme...*, (iv) des marques graphiques d'énumération telles que tirets, puces ou numéros.

- (3) *Les rares compagnies qui le font, et qui modulent leurs tarifs en conséquence, sont unanimes : le risque ne dépend absolument pas de la cylindrée mais, **d'une part**, de « l'image » du modèle concerné (les « sportives » ont plus d'accidents que les routières) **et, d'autre part**, et **surtout**, de l'expérience du pilote. (LMD, oct. 93)*

3.4. Contraintes sur la co-occurrence des introducteurs

Les séries attestées témoignent, nous l'avons vu, d'une grande « souplesse » dans le signalement discursif de leurs composants. Pour autant, cela n'est pas synonyme d'une combinatoire totalement libre entre les introducteurs. Il est donc intéressant de s'interroger sur l'existence d'éventuelles contraintes régissant leur association au sein d'une série particulière. Nous laissons de côté la dimension stylistique dans le signalement des séries. Parmi les questions qui nous concernent plus directement, il y a notamment celle qui touche à la répétition des introducteurs relais qui ne marquent pas la position de l'item (*autre, ensuite, puis...*). Parmi les séries attestées, une seule présente la répétition de balise (*d'abord/puis/et puis*), avec toutefois la présence du connecteur

et devant la seconde occurrence de *puis*. Notons que parmi les séries de quatre items ou plus, une seule recourt uniquement aux introducteurs relais qui ne précèdent pas l'ordre des items. On peut donc en conclure que le phénomène de répétition, sans être impossible, reste réellement marginal et qu'il existe une tendance forte à différencier entre eux les items d'une série, ce qui passe par le choix des introducteurs devant être, eux aussi, différents. Dans le cas des séries longues, le signalement est en plus généralement régulier (une fois ouverte par *premièrement*, la série se poursuit avec *deuxièmement*, *troisièmement*, *quatrièmement*, etc.).

L'existence possible d'emboîtements autorise la succession de deux balises d'ouverture ou de deux balises de clôture. On note alors que les auteurs prennent généralement soin de les différencier nettement entre elles. Dans l'extrait (4), on a ainsi l'introducteur *Le premier* suivi par *Premièrement*.

- (4) *Si l'on tente de formaliser quelques évolutions possibles du continent, ce que l'on pourrait nommer deux « types idéaux » s'imposent. Le premier est celui de la segmentation généralisée. Elle pourrait être précipitée par trois éléments. Premièrement un échec politique et institutionnel à l'Ouest (...) et à l'Est (...). Cette double évolution signifierait (...) Deuxième élément : le maintien d'importantes quantités d'armes de toutes natures sur le continent, alors que s'affaiblissent les pouvoirs ayant vocation à les contrôler. (...) Troisième élément, un développement économique anarchique couplé à une démocratisation chaotique, dessinant en Europe des zones très différenciées : trous noirs, zones d'ombres, monde riche, etc. (...) Dans un continent ainsi haché, (...) L'autre « type idéal » d'évolution de la situation européenne est celui de la recomposition (...). (LMD, décembre 92, pp. 10-11)*

4. NATURE COGNITIVE ET LOGIQUE DE L'OBJET « SÉRIE »

La sériation est l'un des procédés fondamentaux permettant d'articuler des contenus textuels¹⁰ (Vandendorpe, 1999). Plus généralement, il convient de souligner que cette opération cognitive intervient dans plusieurs sphères d'activité humaine. Elle est de toute évidence impliquée dans le processus général de planification d'action. Selon Piaget, elle est nécessaire à l'acquisition du concept de nombre (Piaget, 1974). Des études plus formelles menées en rapport avec la sériation, notamment par Grize (Grize, 1974), mettent en évidence trois propriétés de base que vérifient les éléments d'une série : être disjoints, être équivalents et être ordonnés. Nous nous proposons de vérifier de quelle façon ces propriétés se traduisent dans et par le discours.

4.1. Découper et baliser

Selon la première propriété, les listes (séries) disjointent les éléments. Dans le discours, une série implique discontinuité et non continuité. Elle trace dans le matériau textuel se déroulant linéairement les frontières des unités, permettant ainsi leur identification et leur différenciation. Les limites de ces segments

10. Comme le montrent des études en ethnologie, les listes (séries, énumérations), sont des objets aussi anciens que l'écriture (Goody, 1979).

doivent être clairement signalées. Les relations syntaxiques et la ponctuation, qui mettent en évidence des unités infra-phrastiques, au-delà de la phrase se voient relayées par la présence de marques discursives, sorte de « balises », dotées de capacités d'indexation et d'articulation. Ces balises ont pour effet de séparer les éléments du discours les uns par rapport aux autres et de les séparer du contexte plus large, dans lequel s'inscrit l'ensemble de la série. Cette séparation, à son tour, aide à interpréter correctement des contenus présentant un risque d'ambiguïté.

En effet, c'est quand le découpage du contenu en unités de sens ne va pas de soi que la présence de balises s'impose. C'est le cas dans l'exemple (5), où le journaliste qui rapporte un extrait d'un livre (dans lequel il a d'ailleurs opéré des coupures) est obligé d'ajouter des balises (qu'il place entre crochets) pour que le lecteur puisse identifier sans ambiguïté les fragments correspondant à chacune des six erreurs dont traite le texte.

- (5) *Jean-Marie Messier livre sa version de la crise de Vivendi Universal – VERBATIM « Claude Bébéar est là, en véritable coprésident » EXTRAITS de Mon vrai journal de Jean-Marie Messier (éd Balland). Les six erreurs de Jean-Marie Messier. Ma première erreur est certainement de ne pas avoir opéré plus vite et plus radicalement la séparation définitive entre Vivendi Environnement [les activités historiques de la Générale des eaux] et Vivendi Universal (...). La deuxième erreur a sans doute été de ne pas faire évoluer le conseil d'administration (...) [Troisième erreur] j'ai trop tardé avant de changer l'équipe Canal+ et relever Pierre Lescure de sa responsabilité (...). J'ai eu tort de me laisser séduire par l'homme et intimider par l'icône (...). Son incapacité à organiser Canal+ était patente (...). [Quatrième erreur] Nous en avons peut-être fait un peu trop, trop vite (...). Les acquisitions de Maroc Telecom ou celle de mp3.com étaient certainement moins prioritaires (...). [Cinquième erreur], la relation avec les politiques français. [Sixième erreur] Ma surcommunication. Je n'ai pas écouté les mises en garde répétées de mes collaborateurs, de mes amis, d'Antoinette [sa femme]. (Le Monde, 12 nov. 2002, page 12)*

Les exemples (6) et (4) illustrent deux autres cas où l'introduction de balises s'impose. Dans l'extrait (6), il s'agit de partitionner en deux un ensemble de plus de 2 éléments. Dans l'extrait (4) (§ 3.4), on note des énumérations enchâssées. Le premier élément d'une série de deux contient à son tour une série de trois items. L'ensemble étant linéaire, il faut des balises explicites pour que le lecteur puisse reconstituer cette structure enchâssée.

- (6) *Il est vrai que, sur l'échelle du social-libéralisme, l'Allemagne se situe quelque part au milieu, entre la France de Lionel Jospin d'une part, l'Espagne de José Maria Aznar et l'Italie de Silvio Berlusconi, d'autre part. (Le Monde, 11 04 02, p. 7)*

4.2. Grouper

La deuxième propriété des listes met en lumière l'équivalence de leurs éléments, lesquels, de ce fait, sont considérés comme étant de même nature ou de même importance. Les items d'une liste sont équivalents, car (i) ils sont regroupés sur la base d'une propriété ou d'un ensemble de propriétés qu'ils partagent tous, (ii) leurs caractéristiques propres ne sont pas considérées. En effet, l'abstraction des qualités qui dégage le principe fédérateur conduit à écarter

les qualités différentielles. Dans le discours, c'est l'amorce qui typiquement exprime le principe qui fédère les items, voir l'exemple (7). Un certain nombre de classificateurs (*cause, élément, cas, axe, fonction, domaine, point, piste...*), dont le trait commun est leur capacité à exprimer des relations d'appartenance (ensembliste ou mérologique), peuvent participer à l'expression de ce principe¹¹.

Il n'existe pas toujours d'amorce explicite. La séquence textuelle qui précède la série peut être interprétée *a posteriori* comme une amorce, si les items de la série entretiennent tous une relation argumentative avec cette séquence. Dans l'extrait (8), une relation d'étayage est instaurée entre une thèse et les différents arguments chargés de l'étayer. L'indice fort d'une telle série « argumentative » est la présence de connecteurs (*parce que, en raison de...*) qui doublent ou remplacent les MIL. L'information sur ce qui unit la série peut également être exprimée par une rétro-évaluation, comme le montre l'exemple (9). Ce segment, placé après les items, conclut en quelque sorte l'ensemble, ou affirme quelque chose de général à propos de la totalité des items.

- (7) *Une fois passées les élections françaises et allemandes, à l'automne, l'Europe va devoir faire six ensembles de choix des plus importants. **Premièrement**, ce sera, au mois de décembre à Copenhague, l'élargissement à vingt-cinq pays. **Deuxièmement**, dans les six premiers mois de 2003, la présentation par la Convention, dirigée par Valéry Giscard d'Estaing, d'une nouvelle Constitution. **Troisièmement**, (...). **Quatrièmement**, (...). **Cinquièmement**, (...). **Enfin**, (...).* (Le Monde, 30 juillet 2002, p. 12)
- (8) *Ayant, à plusieurs reprises, exprimé son admiration pour le général putschiste chilien Augusto Pinochet, le général Lebed serait-il tenté par le coup d'État ? Contrairement à ce qu'a prétendu, le 16 octobre dernier, le ministre de l'intérieur, M. Anatoli Koulikov, c'est fort improbable. **En premier lieu**, parce qu'Alexandre Lebed connaît l'état de profonde désorganisation des forces armées ; **ensuite**, parce qu'une tentative de cette nature entraînerait la dislocation de la Fédération de Russie de la même manière que la tentative de putsch de 1991 entraîna la dislocation de l'Union soviétique ; **enfin**, parce que le général, qui est un « national-populiste pur », perdrait toute sa popularité. Car, si la société réclame effectivement de l'ordre, c'est un ordre fondé sur la justice et ne portant pas atteinte aux libertés.* (LMD, novembre 1996, p. 1)
- (9) *Imaginons le téléspectateur de base, si cette catégorie, non répertoriée par l'Insee, existe encore. Qu'est-ce qu'il en déduit le téléspectateur de base ? **Premièrement**, que M. Sarkozy n'habite pas Saint-Germain-des-Prés. Exact, il est domicilié à Neuilly. **Deuxio**, qu'il n'est pas content qu'on montre à la télé, reportages à l'appui, que rien n'a changé sur le terrain depuis qu'il est là, et que la situation a même empiré. **Troisièmement**, qu'Elise Lucet l'énerve un peu, beaucoup, passionnément. **Bref**, on avait droit, jeudi soir sur France 3, à une de ces empoignades entre ministre et journaliste comme on en voit peu sur le PAF.* (Le Monde, 19 octobre 2002, p. 34)

11. Notons que les classificateurs « temporels » (*temps, phase, étape, épisode...*) sont un indice fort en faveur d'une organisation temporelle des contenus « sériés » et non uniquement énonciative, comme c'est le cas pour la plupart des séries introduites par des MIL. Le cas des séries temporelles mérite de toute évidence une étude en soi.

4.3. Ordonner

La troisième propriété d'une liste (série), intimement liée aux deux précédentes, veut que ses éléments soient ordonnés. Étant donné que tous les éléments d'une série sont équivalents, la seule façon de les distinguer est de les ordonner et de leur associer une position. L'ordre d'énumération est donc la seule propriété compatible avec la propriété d'équivalence. Dans le discours, les items d'une liste sont forcément ordonnés, en suivant le flux de la chaîne verbale. Mais le fait de donner une indication supplémentaire, qu'elle soit relative (*l'autre, d'un autre côté, le dernier...*) ou absolue (*le deuxième, secondement, dans un deuxième temps...*), sur la position d'un item, permet de l'isoler plus facilement de l'ensemble, de l'identifier parmi les autres items pour y faire référence. Dans les séries longues (c'est-à-dire celles dont la longueur est égale ou supérieure à quatre items), les introducteurs précisent en règle générale le rang de l'item (*en troisième lieu, quatrièmement,...*). Dans notre corpus, une seule série (sur 91) échappe à cette règle : {*D'abord, puis, ensuite, enfin*}.

5. OPÉRATIONS, STRUCTURES ET CONFIGURATIONS DE MARQUES

Les observations que nous venons de livrer montrent que la sériation dans le discours est bien une opération complexe faisant appel à la fois au balisage des unités textuelles, à leur groupement et à leur mise en ordre. Mais la nature discursive des items (et de leurs balises) intègre également une autre dimension qui, elle, est proprement langagière. En effet, sur les opérations de caractère purement organisationnel, peuvent se greffer des relations argumentatives d'étaillage s'établissant entre le segment amorce et les items de la série, ou, dans le cas des séries binaires, des rapports sémantiques d'opposition, de concession, de hiérarchisation ou conséquence entre leurs items. Les marques de ces relations sémantiques ou argumentatives peuvent alors endosser également un rôle d'organisateur discursifs.

La mise en évidence de ces différentes opérations discursives constitutives de la sériation (ou compatibles avec cette structure) permet de dégager plusieurs structures apparentées entre elles, mais sémantiquement et morphologiquement différentes. À chaque type de structure correspond un ensemble différent d'opérations, ce qui se traduit par une configuration de marques différente. C'est seulement à l'intérieur d'une configuration de marques propre à une structure particulière qu'il est possible de distinguer des indices déclencheurs et des indices complémentaires, comme le préconise la méthode de l'exploration contextuelle que nous avons adoptée pour formaliser nos données. Rappelons que cette méthode consiste à identifier dans un texte, en premier lieu, les unités linguistiques appelées *indices pertinents* ou *indicateurs* qui sont significatifs pour la caractérisation (étiquetage) des segments en fonction d'une tâche donnée, et ensuite de rechercher, avec des règles spécifiques,

des *indices complémentaires* dans le contexte des indices pertinents, indices complémentaires qui vont permettre la prise de la décision adéquate.

| Structure | Opération | Configuration de marques linguistiques | | |
|---|-------------------------------|--|--|---|
| | | Indicateur | Indices complémentaires | Contraintes |
| 1 : série avec amorce et/ou rétro-évaluation | groupement, balisage, ordre | amorce (classificateur, quantificateur) | MIL _i | <i>présence d'au moins un MIL</i> |
| | | MIL _i | reformulation ou quantification | <i>présence d'au moins un MIL</i> |
| 2 : série « argumentative » | balisage, ordre (étayage) | MIL _i | connecteurs argumentatifs | <i>présence d'au moins un MIL et d'au moins deux marques argumentatives</i> |
| 3 : série amorcée et exemplifiée par un seul item | groupement, balisage | MIL ₁ : {en tout premier lieu...} | amorce ; absence de MIL _i (i ≥ 2) | |
| 4 : binaire | balisage (rapport sémantique) | MIL ₁ | absence d'amorce ; MIL ₂ et/ou marque de rapport sémantique | |

Tableau 2 : Les séries : structures, opérations et marques sous-jacentes.

Comme le montre le tableau 2, nous avons distingué quatre types de structure en série.

La première, prototypique, fondée sur la composition des opérations de balisage, de groupement et d'ordre, présente sur le plan de la signalisation linguistique deux variantes. Dans le cas où la série contient une amorce explicite, c'est précisément cet élément qui constitue l'indice déclencheur, les MIL apparaissant comme des indices complémentaires, non tous nécessaires. Dans le cas où il n'existe pas d'amorce, c'est l'occurrence du MIL de rang un ou deux qui constitue le déclencheur, les autres MIL et les marques de rétro-évaluation ayant le statut d'indices complémentaires. La deuxième structure, qualifiée d'argumentative, a ceci de particulier qu'elle ne contient pas d'amorce et que chaque item est précédé (en plus d'un MIL ou à sa place) d'une marque argumentative (*parce que, en raison de*). Le groupement n'est opéré qu'*a posteriori*, car ce qui réunit les items entre eux c'est le fait de former plusieurs arguments distincts en faveur d'une même thèse. Le troisième type de structure en série contient l'amorce et un seul item jugé le plus important ou le plus représentatif. L'indice déclencheur est constitué par l'introducteur du premier item qui le plus souvent est *en tout premier lieu*. L'amorce est l'indice complémentaire important, car c'est elle qui informe de l'existence de plusieurs éléments qui forment effectivement la série. Le deuxième indice complémentaire est l'absence de MIL de rang égal ou supérieur à deux. La quatrième structure, marginale en tant que série mais très fréquente dans les textes, est constituée de

deux éléments nettement différenciés entre eux (opposés, hiérarchisés...) – voir l'exemple (3), § 3.3. Ce n'est donc pas un principe de similitude qui réunit de prime abord ses items, d'où l'absence de relation de groupement. Le rang des items n'étant pas pertinent, seul importe le besoin de les différencier entre eux. Sur le plan de l'expression linguistique, cette structure ne contient ni d'amorce, ni de rétro-évaluation, mais deux balises, dont la deuxième marque une différenciation forte, le plus souvent une opposition (*d'une part – mais, de l'autre...*).

Selon le type particulier de la structure, un élément donné peut jouer des rôles différents. Ainsi, dans la structure [1], c'est l'amorce qui joue le rôle du déclencheur. Dans la structure [3], elle n'est qu'un indice complémentaire, dans la structure [4], c'est son absence qui constitue un indice.

6. CONCLUSION

La tâche d'identification des séries a mis à l'épreuve les principes de la méthode d'exploration contextuelle destinée à formaliser les connaissances linguistiques relatives à cette structure. Nous avons montré qu'il n'était pas possible d'exhiber un seul type de marque comme indice déclencheur pour l'identification des séries. Cette structure admet plusieurs variantes, résultant chacune d'une manière particulière dont se composent les opérations organisationnelles (balisage, groupement, mise en ordre) et certaines opérations sémantiques ou argumentatives (opposition, étayage...). À chaque ensemble d'opérations correspond une configuration de marques particulière, au sein de laquelle on distingue des indices d'emblée saillants et des indices complémentaires. La méthode d'exploration contextuelle se révèle donc pleinement pertinente pour représenter des données linguistiques complexes.

Nos résultats montrent plus généralement qu'il est possible, à condition de disposer d'outils théoriques et informatiques adéquats, d'aborder d'une manière rigoureuse et reproductible l'étude de faits de discours qui échappent *a priori* aux méthodes linguistiques classiques. En effet, chercher à décrire et à modéliser des organisations discursives amène le linguiste, confronté à des données textuelles complexes et nécessairement volumineuses, à modifier en profondeur ses méthodes de travail. Analyser des configurations « mouvantes » de marques, témoins des intrications complexes d'opérations discursives multiples, mais non toutes nécessaires, en intégrant de plus la dimension des usages et des normes, n'est possible qu'avec l'aide d'outils informatiques permettant de recueillir des données, de les quantifier et les analyser par des jeux de paramètres appropriés aux phénomènes étudiés. Cet effort préliminaire se solde par la possibilité de formuler et de tester des hypothèses sophistiquées, de mettre en évidence des phénomènes ou des corrélations inattendues. De plus, les ressources linguistiques ainsi capitalisées acquièrent une « autonomie » certaine vis-à-vis à leur auteur. Elles peuvent donc plus facilement être réexaminées à la lumière d'autres théories ou réutilisées dans plusieurs applications effectives. L'étude sur les séries linéaires que nous venons de présenter illustre pleinement, à nos yeux, les caractéristiques d'une

telle recherche. Nous sommes actuellement en train d'appliquer la même démarche à l'étude des discours rapportés où l'intrication complexe des opérations de citation, de reformulation et de prise en charge appelle des ressources et des outils méthodologiques comparables.

Références bibliographiques

- AUTHIER-REVUZ, J. (1995), *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, t.1 et t.2, Larousse.
- BORILLO, A. (1985), « Discours ou méta-discours », in DRLAV, *Métalangue, métadiscours, métacommunication*, n° 32, pp. 47-61.
- CHAROLLES, M. (1997), L'encadrement du discours : Univers, champs, domaines et espaces, *Cahier de Recherche Linguistique* 6, Université de Nancy 2, <http://www.lattice.enrs>.
- COUTO, J., FERRET, O., GRAU, B., HERNANDEZ, N., JACKIEWICZ, A., MINEL, J.-L., PORHIEL, S. (2004) « RÉGAL, un système pour la visualisation sélective de documents », in *Revue d'Intelligence Artificielle*, vol. 18, n° 4/2004, pp. 481-514.
- CRISPINO, G., JACKIEWICZ, A., MINEL J.-L. (2004), « Spécification et implantation informatique d'un langage de description de structures discursives », *TALN 2004*, Fès, Maroc.
- DESCLÉS, J.-P. (1997), « Systèmes d'exploration contextuelle », in *Co-texte et calcul du sens*, Claude Guimier (éd.), Presses Universitaires de Caen, Caen, pp. 215-232.
- DESCLÉS J.-P. et GUENTCHÉVA Z. (2000), « Énonciateur, locuteur, médiateur », in A. Becquelin et Ph. Erikson (éds), *Les rituels du dialogue*, Éditions de l'Hamattan.
- GOODY, J. (1979), *La raison graphique. La domestication de la pensée sauvage*. Les Éditions de Minuit, traduit de l'anglais et présenté par J. Bazin et A. Bensa.
- GREVISSE, M. (1993), *Le bon usage*, De Boeck & Duculot, 13^e édition.
- GRIZE, B. (1974), « Du groupement au nombre » in *Problèmes de la construction du nombre, Études d'épistémologie génétique*, n°XI, pp. 69-96, K. Reprint, Nendeln/Lienchtenstein.
- JACKIEWICZ, A. (2002), « Repérage et délimitation des cadres organisationnels pour la segmentation automatique des textes », *CIFT'02*, Hammamet, Tunisie, pp. 95-107.
- JACKIEWICZ, A. et MINEL, J.-L. (2003), « L'identification des structures discursives engendrées par les cadres organisationnels », *TALN 2003*, Batz sur Mer.
- Le Monde* (2002), *Le style du Monde*, Guide rédactionnel.
- LUC C. et VIRBEL J. (2001) « Le modèle d'architecture textuelle : fondements et expérimentation », *Verbum*, 23 (1), 103-123.
- MINEL, J.-L., CARTIER, E., CRISPINO, G., DESCLÉS, J.-P., BEN HAZEZ, S., JACKIEWICZ, A. (2001), Résumé automatique par filtrage sémantique d'informations dans des textes, Présentation de la plate-forme FilText, in *Technique et Science Informatiques*, n° 3, PARIS, pp. 369-396.
- PÉRY-WOODLEY, M.-P. (2000), *Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle, Mémoire d'habilitation*, Université de Toulouse-le Mirail.
- PIAGET, J. (1974), « La construction du nombre » in *Problèmes de la construction du nombre, Études d'épistémologie génétique*, n°XI, pp. 1-68, K. Reprint, Nendeln/Lienchtenstein.
- REY-DEBOVE, J. (1997), *Le métalangage*, Arman Colin.
- TURCO G. et COLTIER D. (1988), « Des agents doubles de l'organisation textuelle, les marqueurs d'intégration linéaire », in *Pratiques*, n° 57.
- VANDENDORPE, C. (1999), *Du papyrus à l'hypertexte. Essai sur les mutations du texte et de la lecture*, Éditions de la Découverte.