

Présentation
Introduction au volume collectif *Lingvisticæ Investigationes*
Ordre des mots dans la phrase française, positions et topologie
à paraître 2006

Kim Gerdes & Claude Muller.

L'ordre des mots est au cœur de la syntaxe et il est d'abord par excellence un niveau observable. Il reflète les contraintes de dépendance des constituants, leur intégration dans des unités de rang supérieur, mais aussi des phénomènes liés à la hiérarchie communicative, thématization et focalisation. Enfin, les unités linguistiques se placent aussi selon leur poids en termes phoniques et prosodiques: les phénomènes de cliticisation sont maintenant bien connus, qui permettent par exemple dans beaucoup de langues de placer différemment des noms les pronoms personnels. L'objectif du colloque des « Journées de la syntaxe »¹ était d'explorer la question de l'ordre des mots et de l'étudier en tant que tel. Ce volume est issu de cette réunion. On trouvera dans les articles de ce volume, essentiellement à propos du français, mais aussi sur d'autres langues dans une perspective contrastive, des exposés de recherches récentes, la plupart en cours, sur l'analyse linguistique des phénomènes d'ordre. L'analyse linguistique de l'ordre des mots s'est basée sur des principes qui ont varié considérablement au cours du temps: ordre « logique » des grammairiens universalistes du 18ème siècle, ordre « stylistique » dans la mouvance saussurienne selon laquelle l'ordre des mots relève de la « parole » et non de la « langue », ordre essentiellement basé sur la constituance des grammaires chomskyennes, avec des règles de déplacement pour obtenir l'ordre observé lorsqu'il y a distorsion entre la structure de constituants et la structure de surface. Dans cette dernière tradition, on avait tendance à occulter l'ordre des mots en considérant qu'il n'était qu'une image imparfaite de la structure profonde, une image « brouillée »². Cette vision était une vision simplificatrice en ce qu'elle privilégiait deux types d'organisation : avant tout, la hiérarchie syntaxique quelque soit la façon dont on l'exprime, en constituance ou en dépendance ; et en deuxième structure, la structure logique rendant compte de la portée des quantificateurs, au détriment de d'autres organisations : d'autres contraintes sémantiques (cf. l'article de C. Fuchs dans ce volume) et notamment la structure communicative (ou informationnelle), mais aussi la structure prosodique (et les contraintes stylistiques qui y sont liées, comme le balancement des groupes prosodiques ou la mélodie générale de la phrase) ou encore des contraintes syntaxiques propres à une langue (comme le V2 en allemand) mais qui ne semblent pas directement imposé par la hiérarchie (en tout cas tous les travaux du cadre génératif qui ont essayé d'en faire une conséquence ont à notre avis échoué). C'est dans le développement des théories structuralistes que l'ordre des mots a reçu des descriptions autonomes, qui en font un véritable niveau d'analyse, quel que soit le lien de celui-ci avec les fonctions des mots. La description de cet observable a donné lieu à une *topologie* notamment dans deux courants de la recherche linguistique, en Allemagne d'abord, puis au Danemark, mais aussi plus récemment avec l'introduction de la macrosyntaxe dans les recherches sur le français parlé.

¹ *Journées de la syntaxe : Ordre des mots dans la phrase française, positions et topologie*, organisées le 26 et 27 novembre 2004 à Bordeaux par l'ERSS, l'Equipe de recherche en Syntaxe et Sémantique, UMR 5610 du CNRS.

² On parle de brouillage (« scrambling », Ross, 1967) quand la construction d'un arbre de constituants immédiats semble difficile parce que deux gouverneurs intercalent leurs dépendants (Ross J.R.. 1967. *Constraints on Variables in Syntax*, Thèse de Doctorat, MIT, Cambridge, Mass.)

Il est intéressant d'approfondir l'histoire des courants « topologiques », c'est-à-dire des types d'approche qui considèrent l'ordre des mots comme un niveau d'organisation propre qui doit être interfacé avec les autres niveaux de représentation syntaxique. La notion de topologie est basée sur l'idée qu'il existe des places (aussi appelés « champs » ou encore « positions fixes ») dans la construction d'une phrase, places qui sont remplies en fonction de différents critères syntaxiques, incluant la catégorie grammaticale, la dépendance, la hiérarchie communicative, le poids etc. Dans une conception topologique des relations syntagmatiques, « il n'existe pas de connexion inhérente entre ce qui peut occuper un champ topologique et la composition catégorielle. En d'autres termes, il n'est pas nécessaire que les éléments qui sont éligibles pour un champ topologique forment également une classe naturelle pour d'autres aspects. [...] De même, nous rejetons l'idée communément acceptée en GB que les positions sont intrinsèquement limitées à un seul niveau de projection » (Kathol 1995:139³). Avec un exemple de l'allemand, Jacobs (1988:12⁴) explique bien les bases de cette approche : « Même si on accepte la supposition – en premier lieu contre-intuitive – que le verbe fini en première ou seconde position se place dans la position du complément, dans le but d'exprimer de manière élégante les relations grammaticales multiples entre les verbes finis dans ces positions et la conjonction de subordination, [...] il n'est pas obligatoire de présupposer que le verbe fini a la *fonction grammaticale* d'un complément dans les phrases de ce type ».

En fait, l'idée de place fixe est apparue très tôt dans la description de l'allemand : on peut la retracer au moins jusqu'aux travaux de Herling 1821⁵ qui construisait la première théorie globale sur la structure hiérarchique des phrases complexes. En 1886, Erdmann⁶ proposait des listes de places de la phrase en énumérant le type d'éléments que ces places pouvaient contenir, sans pour autant donner des noms spécifiques à ces places. A cause de la complémentarité de ces deux travaux pionniers de ce qu'on appelle aujourd'hui *la théorie des champs topologiques*, Höhle 1986⁷ va même proposer d'appeler cette théorie le système de Herling/Erdmann. Le terme de « champ » pour désigner ces places dans la phrase apparaît pour la première fois chez Drach 1937⁸ dans un livre intitulé *Idées fondamentales de la science de la phrase allemande*⁹ destiné aux enseignants de l'allemand comme langue maternelle et langue étrangère.

Le livre de Drach prône, avec une teinte légèrement nationaliste, une émancipation de la grammaire, qui était jusqu'à là sous une influence latine forte, en faveur d'une « construction d'une présentation et d'un système de règles basé sur la nature de la langue allemande »¹⁰ (Drach 1937:6). Ce qui est particulièrement intéressant dans cet ouvrage, c'est la liaison que l'auteur fait entre la structure de la phrase séparée en champs et la phonétique¹¹. Dans la tradition explicite de Bally et de Taine, Drach énonce : « nous avons besoin d'une science du son de la phrase. La constatation doit s'imposer que les moyens biologiques d'expressions de

³ Kathol, A., 1995. *Linearization-based German Syntax*, Thèse de Doctorat, Ohio State University.

⁴ Jacobs, J. 1988. « Probleme der freien Wortstellung im Deutschen ». *Sprache und Pragmatik*. Arbeitsberichte 5. pp. 8-37

⁵ Herling, S. (1821). « Über die Topik der deutschen Sprache » *Abhandlungen des frankfurterischen Gelehrtenvereins für deutsche Sprache* pp. 296-362.

⁶ Oskar Erdmann. 1886. *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung dargestellt*. Stuttgart.

⁷ Tilman Höhle 1986. « Der Begriff 'Mittelfeld', Anmerkungen über die Theorie der topologischen Felder » *Akten des Siebten Internationalen Germanistenkongresses*, Göttingen, pp. 329-340.

⁸ Drach, E. 1937. *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt.

⁹ *Satzlehre* 'science de la phrase' est le terme allemand, bien que peu usité, de *syntax* 'syntaxe'.

¹⁰ texte original : « Aufbau einer im Wesen der deutschen Sprache begründeten Darstellung und Regelfassung »

¹¹ Ce que Drach n'appelle bien entendu pas phonétique (*Phonetik*) mais *Schall-lehre* (science du son)

la langue vivante parlée – l’accent, l’intonation, les pauses – sont partie essentielle de l’intégralité de la langue »¹². (Drach 1937:8).

Bech 1955¹³, travaillant à Copenhague avec Diderichsen, a adapté par la suite la terminologie de Drach pour la construction de la structure interne du complexe verbal de l’allemand. Griesbach 1960¹⁴ fait la connexion explicite entre la terminologie de Drach et le système de Herling/Erdmann, encore avec des noms légèrement différents. Enfin, ce n’est qu’avec Engel 1970¹⁵ que la terminologie des champs telle qu’elle est utilisée aujourd’hui sera (presque) définitivement établie. Höhle 1986 reviendra sur cette terminologie en remplaçant les noms des champs par des variables abstraites comme X et Y et en introduisant l’idée d’un champ supplémentaire devant le Vorfeld (le « préchamp » devant le verbe fini), une place pour certaines constructions avec un dédoublement de places devant le verbe fini.

Les travaux sur la topologie de l’allemand ont pris un tournant formel avec les travaux de Reape 1994¹⁶ et Kathol 1995 qui utilise le modèle topologique dans une grammaire syntagmatique guidée par les têtes (HPSG). Ils introduisent un trait *DOM* responsable de la linéarisation des mots et ils enlèvent ainsi la tâche de l’ordonnance à la structure de constituance, sur laquelle se base HPSG. Dans ces approches, la structure de constituance exprime donc surtout la hiérarchie dépendentielle des mots. Plus loin dans cette même direction vont les travaux formels en grammaires de dépendances (Bröker 1998¹⁷, Debusmann & Duchier 2001¹⁸, Gerdes & Kahane 2001¹⁹) : la topologie devient un système de contraintes permettant de linéariser aussi des structures non-projectives, c’est à dire les structures où un mot ne se place pas par rapport à la place de son gouverneur syntaxique mais par rapport à un autre mot de la phrase (comme par exemple les clitiques du français qui se placent par rapport à l’auxiliaire et pas en fonction du participe passé même si ce dernier les régit syntaxiquement). Les descriptions en grammaires de dépendance deviennent ainsi capable de traiter élégamment ce qu’on appelle « les dépendances de longue distance ».

Au Danemark, c’est dans les discussions du Cercle de linguistique de Copenhague que s’est développée une analyse topologique, sous le terme d’analyse « topique », essentiellement avec les contributions de Poul Diderichsen²⁰ et dans une moindre mesure du romaniste Togeby²¹. Cette réflexion s’appuie sur des travaux antérieurs de Drach dont Diderichsen reprend l’idée d’une tripartition en trois « champs » principaux, et la raffine en subdivisant ces

12 Texte original : « Notwendig ist eine Schall-Lehre des Satzes. Die Erkenntnis muss sich durchsetzen, dass die biologischen Ausdrucksmittel lebendigen Sprechens – die Betonung, die Intonation, die Einschnitte – als notwendiger Bestand mit zum Sprachganzen gehören. »

¹³ Bech, G. 1955. *Studien über das deutsche Verbum infinitum*, 2e édition 1983, Linguistische Arbeiten, Nr. 139, Niemeyer, Tübingen.

¹⁴ Heinz Griesbach 1960 *Neue Grundlagen für den fortgeschrittenen Deutschunterricht I, II, III, IV. – Deutschunterricht für Ausländer.*

¹⁵ Ulrich Engel. 1970. *Regeln zur Wortstellung. Forschungsberichte des Instituts für deutsche Sprache 5*, Institut für deutsche Sprache, Mannheim.

¹⁶ Reape, M., 1994. « Domain Union and Word Order Variation in German », J. Nerbonne et al. (éds.), *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes, N° 46, Stanford.

¹⁷ Bröker, N. 1998. « Separating Surface Order and Syntactic Relations in a Dependency Grammar ». Actes de *COLING-ACL 98*.

¹⁸ Duchier, D., R. Debusmann 2001. « Topological Dependency Trees: A Constraint-based Account of Linear Precedence », dans *Proceedings ACL 2001*, Toulouse.

¹⁹ Gerdes, K., S. Kahane 2001a. « Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy », *Actes d’ACL 2001*, Toulouse.

²⁰ Les linguistes danois renvoient aux textes de Diderichsen en danois (cf. Herslund dans ce volume). On trouvera des exposés de Diderichsen en anglais ou en allemand dans le Bulletin du Cercle Linguistique de Copenhague, années 1941-1965 (Akademisk Forlag, Copenhague, 1970), dans quelques pages de sa grammaire anglaise: P. Diderichsen, 1964, *Essentials of English Grammar* (5th issue, 1982, Akademisk Forlag, Copenhague), et dans le recueil d’articles : P. Diderichsen, 1976, *Ganzheit und Struktur*, Wilhelm Fink Verlag, München. Cette analyse ainsi que celle de Togeby est résumée dans: Magali Rouquier, 2004, *Réflexions sur l’Analyse en Zones de la proposition et l’Approche Pronominale*, *Travaux de Linguistique*, 48, pp. 53-75.

²¹ Un compte rendu de l’analyse de Togeby figure dans le Bulletin du Cercle Linguistique de Copenhague (1970: années 1941-1965) à la date du 11-12-1947.

champs en « places ». La méthode consiste à dégager une séquence maximale de constituants, chacun affecté de sa « signature » (le type auquel appartient le constituant). Il peut y avoir plusieurs séquences maximales, donc plusieurs types de base; les formules plus courtes sont classées en utilisant les similitudes de signatures pour préserver l'ordre des éléments, ce qui conduit à laisser des blancs dans leur présentation. Les champs (la tradition en français utilise plutôt le terme « zone » -Skårup 1975) sont des regroupements d'unités que l'on distingue par leur position linéaire et non par leur fonction dans la phrase (par exemple, des séquences de constituants qui précèdent toujours le verbe seront placés dans un champ initial).

Parmi les éléments de la signature, il y a la catégorie grammaticale, la fonction, mais aussi le poids. Ce critère est expressément utilisé (par exemple par Togeby dans son exposé de 1948 au cercle de Linguistique de Copenhague) pour distinguer des éléments normaux des « légers » (par exemple les pronoms personnels, mais aussi certains termes de quantité comme *tout* préposé au participe passé dans *il a tout mangé*) et des « lourds » (par exemple les compléments de type phrastique en allemand, placés en subordonnée après la séquence de verbes finaux et non avant).

Diderichsen a appliqué sa méthode à l'anglais, l'allemand et le danois et la linguistique danoise a poursuivi dans cette voie (cf. Herslund ici même). Povl Skårup²² a utilisé brillamment cette méthode sur l'ancien français, langue ayant une structure assez proche (par ses zones préverbale et verbale) de celle du danois ou de l'allemand moderne. Donohue & Sag 1999²³ ont proposé une analyse basée sur l'idée de place pour le Warlpiri, et Cori et Marandin 1993²⁴ et Marandin 2001²⁵ ont formalisé le terme de « site » de Milner 1989²⁶ pour l'analyse topologique du français moderne.

Ce qui définit les descriptions topologiques, c'est leur caractère strictement statique, l'idée qu'une phrase propose un gabarit de champs dans lesquels les mots se placent. Cependant, l'occupation de la position initiale (celle de la « base » ou du « fondement » dans la terminologie de Skårup) pour des raisons de thémativité dans les langues de type « verbe-second » comme le danois ou l'ancien français conduit à laisser une place vide dans un autre champ, donc donne en quelque sorte l'équivalent d'une analyse par déplacement à gauche d'un constituant laissant une trace dans sa position de base (ce serait ici la place laissée obligatoirement vide). Par la suite, sous l'influence des grammaires chomskyennes, les descriptions statiques de l'école danoise de topologie ont pu incorporer l'idée de placements ou déplacements de constituants, donc celle qu'il y a des structures basiques et des structures dérivées (par exemple dans les constructions à interrogatif initial). On trouve cet infléchissement dans les analyses topologiques danoises (cf. *Copenhagen Studies in Language*, 15, *Word Order*, 1992).

Un autre infléchissement de la topologie tient à l'importance considérable que l'ordre des mots a prise dans les développements de la typologie, à la suite des travaux de J. Greenberg²⁷. L'ordre des mots a cessé alors définitivement d'être la conséquence marginale d'un arrangement stylistique de la constituance, pour devenir un véritable niveau de représentation et d'une certaine façon un élément essentiel de la caractérisation typologique. D'où les études contrastives dont on trouvera des exemples dans ce volume.

Nous allons maintenant donner une description brève du contenu des contributions.

²² Cf. P. Skårup, 1975: *Les premières zones de la proposition en ancien français*, Akademisk Forlag, Copenhague.

²³ C. Donohue, and I. A. Sag. 1999. « Domains in Warlpiri » *Sixth International Conference on HPSG*.

²⁴ Cori M., J.M. Marandin 1993 « Grammaires d'arbres polychromes », *TAL*, volume 34, numéro 1, pp. 101-132.

²⁵ Marandin J.-M. 2001. « L'hypothèse des sites en syntaxe » dans Marandin J.-M. (eds.) *Cahier Jean-Claude Milner*, Verdier, pp. 175-227.

²⁶ J.C. Milner 1989: *Introduction à une science du langage*, Le Seuil, pp. 291 et suivantes.

²⁷ J. Greenberg, 1966: Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements, in: Greenberg, ed.: *Universals of Language*, MIT Press, Cambridge, Mass., pp. 73-113.

Dans leur article, A. Abeillé et D. Godard proposent une contrainte de légèreté: la légèreté bloque la mobilité et suppose des contraintes d'ordre basées sur la tripartition faible / léger / non-léger. On notera avec intérêt que les mots légers peuvent, au contraire des formes faibles, être coordonnées.

Dans son étude de l'inversion nominale locative, A. Borillo décrit la structure informationnelle très particulière de cette construction, avec un verbe porteur de peu d'information, et un sujet qui porte l'information essentielle, le tout étant inséré dans un discours préalable, en quoi l'inversion locative se distingue des tournures existentielles.

Sur le même sujet, les inversions locatives, C. Fuchs montre qu'il n'y a pas de relations de dérivations de l'énoncé à locatif en tête et inversion du sujet à partir de l'énoncé à locatif en tête et sujet préverbal, pas plus qu'il n'y en a à partir de la structure à locatif final. Elle expose ensuite un modèle topologique de construction de l'énoncé basé sur la notion sémantico-cognitive de « repérage », ces repérages étant organisés en cascade.

Paul Boucher s'intéresse à la place des adjectifs épithètes, avant ou après le nom. Il montre que les analyses par Raising sont inappropriées: la place des adjectifs est mieux décrite par des contraintes sémantiques, ce qui l'amène à postuler la présence d'opérateurs logiques d'existence et de généricité dans le syntagme nominal.

K. Gerdes et S. Kahane appliquent au français le modèle topologique (proposé pour l'allemand dans la Théorie Sens-Texte) où un arbre topologique est construit dans la linéarisation d'un arbre de dépendance. L'originalité de leur propos est la notion « d'amas verbal » emprunté à leur analyse de l'allemand. Dans ce modèle, l'inversion du sujet nominal est décrite à partir de la notion d'amas verbal, ce qui explique que le sujet inversé puisse être placé dans les enchâssées. Les auteurs supposent alors que le champ du sujet est occupé, par exemple par un locatif. Ils examinent aussi le niveau macrosyntaxique, qui suppose des interactions entre la structure communicative et les fonctions syntaxiques des mots.

L. Gournay s'intéresse à l'inversion absolue (il n'y a pas de terme précédant le verbe) et montre que cette catégorie d'énoncés a des propriétés spécifiques (type de verbes, emploi des temps, connecteurs), renvoyant à une fonction différente qui relève pour l'auteur de la cohérence discursive.

M. Herslund plaide pour l'autonomie d'un niveau topologique de représentation. Il illustre ce niveau par les trois exemples de la topologie du danois, de l'ancien français, assez semblable et descriptible aussi comme une langue V2, et le français moderne, dont la topologie est très différente. M. Herslund propose de prendre comme pivot de la description le sujet nominal, ce qui conduit à deux positions distinctes du verbe, et à des réalisations très différentes pour le sujet clitique, décrit comme un appendice morphologique du verbe. Sa description se place dans la tradition topologique danoise issue des travaux de Diderichsen.

Pour K. Lahousse, qui s'intéresse aux inversions du sujet nominal en subordonnée, le critère discriminant est le caractère asserté ou non du contenu de la subordonnée. Elle montre dans son travail que les inversions ne sont possibles dans le cas d'un contenu subordonné asserté que si la subordonnée est introduite par un topique. Par contre, les subordonnées à contenu non asserté ne nécessitent aucun terme initial. L'explication réside selon elle dans la possibilité d'interpréter le sujet postverbal comme le topique, lorsqu'une interprétation assertive est possible: la présence explicite d'un autre topique évacue cette interprétation et maintient au sujet postverbal sa fonction communicative de focus.

I. Magnus compare l'ordre des constituants de la proposition indépendante en français et en néerlandais. Les deux organisations sont très différentes, puisque le néerlandais est une langue de type V2, décrite ici avec un système topologique à deux points fixes, le verbe auxiliaire et le verbe principal. Le français, lui, est une langue à ordre fonctionnel. Les compléments périphériques sont comparés dans les deux langues. Il ressort de la comparaison que les propriétés d'un constituant circonstanciel en position initiale, suivi après rupture prosodique de la proposition, sont très différentes dans les deux langues: en néerlandais, il doit y avoir

une reprise anaphorique du terme extraposé, ce qui n'est pas le cas en français. Cette propriété est mise en rapport avec le type V2 du néerlandais.

On dit souvent que l'objet direct ne peut pas débiter une proposition. F. Sabio montre que cela n'est pas vrai, du moins en français parlé, et il distingue deux types très différents d'antéposition: dans l'un, assez contraint, l'objet est un constituant dépendant du noyau propositionnel, traité en utilisant la terminologie développée à Aix comme un « préfixe »; dans l'autre, l'objet est le constituant essentiel de la proposition, en termes communicatifs, et le reste, qui suit, est traité comme un « postfixe » dans la terminologie macrosyntaxique (intonation basse et plate). Les deux types sont donc profondément différents du point de vue informatif.

Toujours dans le cadre de la macrosyntaxe développée à Aix, S. Teston-Bonnard montre que les positions occupées par les constituants non régis obéissent à des critères différents de ceux des régis, la catégorie « non régis » faisant référence à des expressions qui n'ont pas de lien clair avec une hiérarchie dépendancielle.

J. Royauté et P. Sabatier décrivent dans leur article l'ordre des constituants dans les prédications à base nominale, et leur étude montre qu'il y a ici un schéma topologique original à cinq positions, dans lequel deux positions remarquables sont distinguées: une position initiale détachée, dans laquelle un argument du prédicat nominal peut figurer, éventuellement son « sujet » sémantique avec une autre préposition que *de*, et la position finale des compléments argumentaux.

R. Moot et C. Rétoré passent en revue les traitements formels de l'ordre des mots, appliqués aux séquences de clitiques des verbes en français, dans une perspective de traitement automatique. Ils proposent une analyse simple de ces constructions, à partir d'analyses catégorielles multimodales avec un petit nombre de modalités et une combinatoire de celles-ci, pour décrire les combinaisons possibles des clitiques et leur montée ou non sur un verbe auxiliaire.

L'article de C. Muller s'intéresse à un tournant crucial dans l'ordre des constituants majeurs de l'énoncé en français: celui pendant lequel le français passe du type verbe second (comme l'allemand ou le néerlandais actuels) au type fonctionnel SVO actuel. Plus que le placement habituel du sujet en première position, c'est la naissance d'un groupe verbal cohésif qui semble amorcer cette évolution, sensible pendant la période du moyen français (essentiellement au XVe siècle) avec un système V2 réaménagé dans lequel le sujet suit souvent un cluster plus ou moins important fait d'un auxiliaire et de participes ou d'infinitifs, construction qui a survécu jusqu'en français moderne dans certains cas.

Nous espérons que cet ensemble de travaux marquera une étape dans la description de l'ordre des mots en français comme un niveau en soi de représentation, et non plus comme la simple linéarisation d'une structure en constituants déjà ordonnée, comme on l'a trop longtemps décrit alors que des observations cruciales montrent que cette position est intenable, même dans une langue comme le français, à ordre grammatical plutôt que communicatif.