

Marie-José BÉGUELIN, Gilles CORMINBOEUF, Florence LEFEUVRE (éds.), Types d'unités et procédures de segmentation, Limoges, Lambert-Lucas, 2019, pp. 270.

Comme Marie-José BÉGUELIN, Gilles CORMINBOEUF et Florence LEFEUVRE le remarquent dans l'*Avant-propos* (pp. 9-12), *Types d'unités et procédures de segmentation* se veut un aperçu des unités linguistiques en usage par rapport à leur histoire, aux analyses, aux pratiques de segmentation et d'annotation dont elles sont à l'origine, à leur application dans des domaines variés, à savoir la syntaxe, la sémantique, la phraséologie, la structuration textuelle, la didactique du texte, l'analyse du discours et des conversations, le traitement et l'étiquetage de corpus oraux et la psycholinguistique de corpus écrits. Si, en effet, différents courants ont abordé le phénomène de la segmentation des unités linguistiques en l'adaptant aux fins des analyses conduites – tel est le cas du Groupe aixois de recherches en syntaxe, du modèle genevois, du Groupe de Fribourg, mais aussi de recherches sur les phrases averbales, sur les constructions, sur le jet textuel –, la question des unités a retenu l'attention de chercheurs suivant des modèles et adoptant des positions différents, comme il ressort des sections et des chapitres qui composent ce travail, dont l'objectif est entre autres de mettre en lumière la diversité des approches présentées en vue de faire avancer la réflexion collective sur le sujet abordé. Ainsi cet ouvrage est-il structuré autour de trois volets principaux à la fois répartis en chapitres, suivis par la *Bibliographie générale* (pp. 221-238), l'*Index nominum* (pp. 239-244), l'*Index rerum* (pp. 245-256) et les résumés de chaque chapitre en français et en anglais (pp. 257-268).

La première partie du volume, *Enjeux épistémologiques* (pp. 13-82), se compose de quatre chapitres.

Dans le premier chapitre, *De la langue au texte : à la recherche des unités perdues* (pp. 15-32), Pierre LE GOFFIC s'interroge sur le statut théorique des unités linguistiques relativement à la tripartition langue/ discours/ texte, en analysant la manière dont la grammaire a traité ces unités dans le temps et dans le rapport au texte pour ensuite souligner leur rôle dans le cadre de la linguistique de corpus. Le grammairien part de l'opposition, rapportée au langage, entre la puissance et l'acte – le langage en acte se présentant à la fois sous l'aspect du discours ou sous l'aspect du texte – pour souligner que, au-delà de leur hétérogénéité, ces termes sont inséparables, mais également que c'est le texte qui est l'élément le plus variable et le plus sujet à variations. A partir d'un exemple d'échantillon d'exposé réellement produit, il montre qu'il est possible d'y reconstituer le déroulement du discours dans le temps, de constater que le locuteur a une compétence linguistique solide, de mettre en évidence le statut incertain du texte en raison de l'effort du locuteur de transformer en mots son intention de signification et de la transmettre à son interlocuteur afin que celui-ci l'interprète, et, au fond, de s'interroger sur le statut des unités. Pour aborder ce dernier aspect, LE GOFFIC compare le paradigme classique qui va de la langue au discours et qui voit donc dans la proposition la seule unité linguistique possible, le texte n'ayant pas d'existence propre mais étant ramené au discours et s'identifiant avec celui-ci, avec le paradigme moderne, développé à partir du XIXe siècle, qui voit d'abord l'émergence du texte écrit, par l'appui sur la phrase en tant qu'unité appartenant à la fois à la langue et au texte, ensuite du texte oral, depuis la fin du XXe siècle. Dans ce contexte, c'est l'irruption de la technologie qui fait perdre à l'oral sa volatilité et qui permet à la phrase de devenir une unité de la langue et du texte, écrit ou

oral. Par ce survol, l'auteur pose donc la question de la segmentation des textes à la lumière de la distinction écrit/ oral et, dans ses remarques conclusives, il souligne le contenu vide de la phrase, en tant qu'unité de la langue et du texte, tout comme, pour autant, son rôle comme modèle de langue. Quant au discours, il relève les défis posés par l'élaboration d'un véritable modèle de discours ou de performance pouvant rassembler les travaux existants et il laisse enfin ouverte la question de la segmentation des textes en raison du manque d'une procédure adéquate sur le sujet, en dépit des avancées de la technologie.

Le deuxième chapitre, rédigé par Dominique LEGALLOIS, aborde les « comportements » discursifs des constructions grammaticales et la relation grammaire-discours qui leur est sous-jacente (*Les constructions grammaticales comme schémas pré-discursifs*, pp. 33-44). L'étude de la métafonction textuelle des constructions, dont la fonction est la structuration du discours, permet à l'auteur de montrer que la compétence grammaticale et discursive des locuteurs se compose entre autres également d'unités non discrètes n'appartenant ni aux unités phraséologiques ni aux unités lexicalisées. Pour ce faire, il se sert de trois exemples qui indiquent l'imbrication entre grammaire, discours et compétence des locuteurs : l'organisation narrative relevant du schéma narratif préparatoire – désormais disparu –, liant à la fois un motif et l'expression du trinôme datation-circonstance-événement ; le schéma, inhérent à la pratique sociale, de clôture/ rupture, dont est présenté le fonctionnement discursif inhérent issu de la compétence des locuteurs et la complémentarité de plusieurs constructions de niveaux différents ; le *first verb* (emprunté à Sacks (1992)), consistant en un indicateur verbal d'une structure organisationnelle liée à un phénomène d'attente et de projection par lequel émerge une connaissance préétablie et mémorisée, donc préalable dans la compétence des locuteurs. LEGALLOIS relève ainsi que ces formes conventionnelles, illustrées par les trois cas de figure ci-dessus, ont en réalité une dimension *a priori* et sont des préfigurations du discours, souvent schématiques, qui sont appelées à contribuer à l'organisation des discours.

Dans le ch. 3, *De la textualité narrative aux faits syntaxiques dans un écrit scolaire. Peut-on articuler micro- et macro-syntaxe dans une perspective didactique ?*, pp. 45-63), Caroline MASSERON présente une analyse ponctuelle d'écrits scolaires réellement produits dans une perspective de formation des maîtres de français et des enjeux didactiques posés par ce type d'écrits. En distinguant la dimension de la langue de celle du discours, dans l'échantillon choisi, comprenant quatre productions narratives écrites réalisées par des élèves dans les mêmes conditions – en fin de cycle primaire ; même âge ; même consigne autour d'un épisode passé – à Libreville et à Metz, l'auteur examine les unités de segmentation, les rangs de structuration de celles-ci, ainsi que les règles et les indicateurs permettant d'effectuer une segmentation d'un texte complet. Elle souligne ainsi la « fragilité syntaxique » du modèle, très répandu, de la phrase graphique écrite, dominant l'analyse des écrits scolaires, tout comme les « fragilités » relevées au sein des quatre écrits d'enfants. L'auteur fait alors appel à l'influence des unités de rang supérieur et de la rédaction elle-même ; aux phénomènes de redénomination ; au faible apport de la mémoire discursive, tout comme au type de périodes et aux unités « pivots », pour prouver que la segmentation des productions est essentielle pour identifier les mécanismes d'encodage textuel et les erreurs qui en découlent, mais aussi que le recours à une démarche alliant micro- et macro-syntaxe ainsi que l'enseignement des phénomènes de cohésion et de l'enchaînement transphrastique devraient être questionnés dans le cadre d'une perspective didactique renouvelée.

Le dernier chapitre de cette section, *Récurtivité des unités dans les discours : enjeux épistémologiques et sémantico-pragmatiques* (pp. 65-82), rédigé par Annie KUYUMCUYAN, aborde, à partir de l'approche de Benveniste (1966), la constitution de l'unité de discours et les reconfigurations éventuelles que celle-ci reçoit par le biais de ce traitement. C'est au sein du cadre dialogal, à l'appui d'exemples tirés de dialogues apparaissant au sein de textes littéraires et d'une approche comparant le modèle hiérarchique de Roulet (2001) avec celui du Groupe de Fribourg (2012), que l'attention de l'auteur est dans un premier temps focalisée sur l'intervention en tant qu'unité de dialogue minimale appliquée à des enchaînements dialogaux inférieurs au format

propositionnel – tirés de Diderot et de Simenon et au sein desquels l'enchaînement dialogal est opposé à l'enchaînement monologal –, composés de propositions incomplètes, pour souligner que l'analyse syntaxique de ces répliques partielles est soumise aux tours de parole des interlocuteurs. Il s'ensuit une comparaison avec un dialogue narratif complet, tiré d'une nouvelle de Mérimée, dont l'analyse peut donner lieu, selon KUYUMCUYAN, à deux lectures distinctes selon le processus dynamique d'interprétation du segment discursif examiné de la part du lecteur, qui relèvera soit d'une requête soit d'une question. Celui-ci, en effet, en segmentant le fragment discursif, hiérarchisera les unités obtenues – à ce propos, l'auteur propose des représentations arborescentes – en vue de placer certaines unités sous la dépendance d'autres unités, créant ainsi une recombinaison potentiellement indéfinie des unités et de leur combinaison.

La deuxième partie du volume, *Modèles d'analyse et choix de segmentation* (pp. 83-162), compte également quatre chapitres.

Dans le chapitre 5, *Clauses nominales : prédication ou monstration ?* (pp. 84-98), pour étudier les clauses nominales « monorèmes » et leurs retombées sur la théorie générale des unités linguistiques, Alain BERRENDONNER ne s'appuie pas sur la conception vérifonctionnelle du langage, héritée d'Aristote, mais sur une sémantique nominale généralisée. Ainsi montre-t-il que ni l'interprétation prédicative ni la modalité énonciative ne sont des invariants ou des propriétés constantes des syntagmes nominaux. C'est pourquoi il est nécessaire, d'après l'auteur, de changer de perspective pour examiner les interprétations contextuelles des clauses nominales, en adoptant une conception du langage comme activité ostensive-inférentielle dont les fondements reposent sur l'idée que la communication verbale se construit autour d'un dialogue *in praesentia* et concerne donc une activité gestuelle de monstration, et que le langage n'est pas une activité verifonctionnelle mais fictionnelle. Dans ce cadre, l'auteur relève que le syntagme nominal – le modèle nominal est généralisé à tous les énoncés – est l'acte ostensif d'un nom, dont l'acte de langage est la désignation d'un objet-de-discours, et que divers effets peuvent être inférés sur la mémoire discursive en termes de localisation, d'apport ultérieur d'informations, de contenu.

Pour sa part, le ch. 6, rédigé par Marie-José BÉGUELIN et Gilles CORMINBOEUF, porte sur les greffes et les segments flottants, à savoir deux types de séquences posant des problèmes de segmentation – leur place syntaxique et leur contenu sémantique – et de délimitation, et d'unités syntaxiques et discursives pertinentes, respectivement (*Segmentation en unités : le cas des « greffes » et des « segments flottants »*, pp. 99-129). A partir d'extraits tirés de la base de français oral Ofrom et d'exemples de français écrit (Frantext), les auteurs distinguent ces segments d'autres structurations apparentées mais distinctes de ceux-ci. Si les greffes d'une construction verbale sont examinées quant à leur position syntaxique – à la suite du verbe ; sous une rection prépositionnelle ; en position adnominale ; en position nominale, après un déterminant ; en coordination –, au plan sémantique, les constructions verbales greffées peuvent soit signaler des circonstances spatio-temporelles du procès, la manière, ou un dire un cours d'élaboration, soit véhiculer un contenu argumentatif ou épistémique, soit concourir à identifier des référents et/ou enrichir leurs propriétés. Pour ce qui est de la segmentation et du traitement de ces constructions, les auteurs comparent les deux modèles aixois et fribourgeois en vue de proposer une typologie des greffes à partir d'une modélisation dynamique de ce phénomène. Relativement aux segments flottants, organisés autour d'une structure ternaire (ABC), à l'appui d'exemples et des cas de figure qui y apparaissent, les auteurs confirment la parenté de ces segments avec la notion grecque d'*apo koïnu* en identifiant quatre configurations pragma-syntaxiques caractérisées par une rupture et une reprogrammation ; un couplage de deux structures attributives ; des « îles flottantes » ; des syllepses syntagmatiques, respectivement. Il émerge non seulement que le terme B n'a pas de statut syntaxique homogène, mais également que les quatre cas de figure mettent en évidence une adaptation du locuteur aux besoins immédiats et entraînent une réanalyse fonctionnelle instantanée. Il émerge par ailleurs que les greffes et les segments flottants, loin d'être des phénomènes typiques du français parlé, sont à bien différencier d'autres phénomènes ; que le contexte métanalytique joue un rôle central dans le

cas des greffes ; que les notions de « mémoire discursive », de routines périodiques et de « métanalyse » en usage, empruntées au modèle fribourgeois, contribuent à l'analyse de ces segments. Dans le ch. 7, *La configuration discursive unité résomptive / unité prédicative c'est vrai, P du type c'est vrai, je t'ai un peu oublié* (pp. 131-147), Florence LEFEUVRE examine le schéma *c'est vrai, P* et ses deux variantes *P, c'est vrai* et *c'est vrai* au sein de *P* à partir des deux unités prédicatives qui le composent, dont l'une, exemplifiée par *je t'ai un peu oublié pendant ce bienheureux printemps* (Frantext), est autonome et l'autre, *c'est vrai*, joue le rôle d'unité résomptive caractérisant ou modalisant la première unité. L'approche théorique suivie par l'auteur repose sur Le Goffic (1993 ; 2011), à savoir sur une conception de la phrase ou des unités discursives à l'interface de la syntaxe et du discours. L'attention est d'abord focalisée sur le lien sémantique entre l'unité résomptive et l'unité prédicative, ce qui montre que ces deux unités représentent un tout sur le plan sémantique. L'analyse se poursuit ensuite par l'étude des contraintes syntaxiques de *c'est vrai* : cette unité est syntaxiquement dépendante de l'unité prédicative, mais l'ensemble de ces unités constitue, selon LEFEUVRE, une « période discursive », c'est-à-dire une unité supérieure constitutive du discours pourvue de valeurs discursives au niveau argumentatif. Celles-ci donnent lieu à trois moments argumentatifs principaux : le renforcement ; la justification ou l'exemplification ; la restriction.

Le dernier chapitre de la deuxième section, rédigé par Georgeta CISLARU et Thierry OLIVE, porte sur des phénomènes d'« amorçage » dans le cadre du processus de textualisation, notamment dans un processus de rédaction enregistré en temps réel (*Dynamiques d'amorçage au cours du processus de textualisation dans l'écriture enregistrée*, pp. 149-162). En tant que processus d'actualisation d'une unité linguistique activant une attente, l'amorçage est étudié par les auteurs en termes linguistiques et psycholinguistiques, relativement aux types de catégories grammaticales et de relations sémantiques, en particulier par rapport aux jets textuels – en examinant leurs frontières et plus en particulier leur borne droite –, pour catégoriser les phénomènes d'amorçage figurant dans le corpus examiné, traité dans la deuxième partie de l'étude. Celui-ci, composé de rapports éducatifs et de dossiers académiques dont le processus de rédaction a été enregistré en temps réel par le biais des logiciels Inputlog et Scriptlog, vise à examiner une sélection de groupes de jets textuels en fonction de leur borne droite et des suites projetées. Trois niveaux de saillance d'amorçage y sont identifiés : si les attentes de nature syntaxique sont soumises à des relations de dépendance et à la combinatoire syntagmatique, les attentes lexico-sémantiques relèvent d'un cadre collocationnel, tandis que les attentes sémantico-fonctionnelles sont à relier à des macrostructures discursives. Il s'ensuit, en termes morpho-syntaxiques, que les catégories des noms, des adjectifs et des verbes sont plus utilisées comme bornes droites projetant une suite dans le processus de textualisation que celles des déterminants et des prépositions : cela est dû, selon les auteurs, aux relations informationnelles et discursives par lesquelles le thème est en attente de spécification et une nouvelle prospection temporelle s'ouvre pour chaque nouveau jet textuel, dont les pauses marquent la frontière entre des séquences à portée générale structurante amorçant une caractérisation ou une spécification.

La troisième et dernière partie du volume, *Analyse sur corpus et/ou segmentation outillée* (pp. 163-220), est composée de trois chapitres portant sur la langue orale.

Dans le ch. 9, consacré à l'espagnol, Oscar Garcia MARCHENA analyse des corpus oraux à partir du corpus CORLEC représentatif de l'espagnol oral contemporain et les défis qu'ils posent par rapport au concept syntaxique de la phrase comme structure construite autour d'un prédicat verbal et de son sujet (*Les fragments comme unités linguistiques : une analyse de corpus de l'espagnol oral*, pp. 165-180). L'attention de l'auteur est en particulier focalisée sur deux types d'énoncés sans verbe, c'est-à-dire les phrases averbales et les fragments, dont il met en évidence les difficultés de classement, les sous-types, les propriétés syntaxiques, sémantiques et illocutoires, et la distribution au sein du corpus en vue d'envisager leurs réalisations sous forme d'énoncés différant de la phrase à tête verbale.

Le français parlé fait en revanche l'objet de l'étude de Mathilde CARNOL et Anne Catherine SIMON (*Forme et fréquence des constructions verbales en français parlé*, pp. 181-201) qui, s'inspirant des travaux de Blanche-Benveniste autour de la syntaxe du français parlé dans le cadre de la grammaire de dépendance, examine le verbe dans la syntaxe de dépendance en tant qu'unité de rection en vue de vérifier la fréquence des unités de rection construites autour d'un verbe et de ses compléments au sein du corpus de français parlé multigenre Locas-F. L'analyse conduite par les auteurs, dans laquelle elles présentent un inventaire des variantes de dispositifs de rection identifiés – dispositif direct ; clivé ; pseudo-clivé ; binarisé ; construction présentative – et la difficulté de classer le dispositif concerné – en témoignent les constructions segmentées réparties en double marquage à gauche et à droite ; en constructions en *A c'est B* et *c'est B A* ; en associé lexical –, vise ainsi à établir non seulement la fréquence de ces dispositifs, mais également les types de compléments qui y apparaissent davantage. Il émerge une vision générale du fonctionnement de ces dispositifs en français parlé contemporain, tout comme des remarques qualitatives sur des structures dont l'annotation et l'analyse posent problème.

Enfin, le ch. 11, *La syntaxe en empirie et en théorie. La proposition de segmentation multiniveau du projet SegCor pour le français parlé* (pp. 203-220), relève de l'analyse de l'oral en interaction dans le cadre du projet SegCor – visant à réaliser une segmentation multiniveau de corpus oraux du français en termes syntaxiques – dans le but d'examiner les choix de rédaction de protocoles d'annotation à différents niveaux syntaxiques. Nathalie ROSSI-GENSANE, Biagio URSI, Iris ESHKOL-TARAVELLA et Maria SKROVEC examinent les types de segmentations proposés dans les protocoles de rédaction en comparant le cadre théorique relevant des études de Blanche-Benveniste et du Groupe de Fribourg en micro- et macro-syntaxe, d'une part, et les projets Rhapsodie et Orfeo, d'autre part, pour souligner que l'appui sur une rection étendue permet d'élargir le domaine de la (micro-)syntaxe au détriment de celui de la macro-syntaxe. Ainsi, étant donné les problèmes de délimitation entre les niveaux d'analyse micro-et macro-syntaxique, les auteurs proposent une segmentation syntaxique en *chunks* – des unités d'analyse minimale syntaxique définies en termes de constituance – dont ils soulignent les décisions opérées par rapport aux étiquettes choisies pour indiquer les locutions figées, les mots composés par plusieurs tokens, les contextes d'occurrence d'une même forme, les disfluences de la parole – et présentent l'unité maximale (micro-)syntaxique et l'unité maximale macro-syntaxique au sein du projet SegCor. Dans leurs remarques conclusives, ils présentent une réflexion sur les unités maximales complexes, aux niveaux micro- et macro-syntaxique, qui comportent des segments apparemment non-dépendants d'un point de vue syntaxique et qui peuvent être considérés comme juxtaposés ou coordonnés.