Linx

Revue des linguistes de l'université Paris X Nanterre

82 | 2021

Entre vieillissement et innovation : le changement linguistique

---

Compte rendu de lecture

Un ouvrage de Marie-José Béguelin, Gilles Corminboeuf et Florence Lefeuvre (direction), *Types d'unités et procédures de segmentation*, Limoges, Lambert-Lucas, 2020, 270 pages.

Hélène Vassiliadou

Édition électronique

URL : <https://journals.openedition.org/linx/7983>

DOI : 10.4000/linx.7983

ISSN : 2118-9692

Éditeur

Presses universitaires de Paris Nanterre

Référence électronique

Hélène Vassiliadou, « Compte rendu de lecture », *Linx* [En ligne], 82 | 2021, mis en ligne le 30 juin 2021, consulté le 20 juillet 2021. URL : <http://journals.openedition.org/linx/7983> ; DOI : <https://doi.org/10.4000/linx.7983>

Ce document a été généré automatiquement le 20 juillet 2021.

Département de Sciences du langage, Université Paris Ouest

Compte rendu de lecture

Un ouvrage de Marie-José Béguelin, Gilles Corminboeuf et Florence Lefeuvre (direction), *Types d'unités et procédures de segmentation*, Limoges, Lambert-Lucas, 2020, 270 pages.

Hélène Vassiliadou

- 1 La segmentation de la chaîne linguistique, intimement liée à des concepts dont le statut théorique et les critères de reconnaissance sont diversifiés (phrase, proposition, énoncé, clause, période, période prosodique, unité illocutoire, construction, chunk, fragment, (pré/post-)noyau, tour de parole, unité intonative, unité d'alignement, syntagme, etc.), est un sujet qui revient régulièrement dans la littérature et qui suscite toujours de nouvelles discussions sur les différents modèles (fribourgeois, genevois, aixois pour l'essentiel) et les domaines d'application. La mise à disposition d'outils et de ressources pour le traitement de l'oral mais aussi pour les écrits dits *peu planifiés* ou pour les manuscrits anciens participent également du renouveau des données, des approches, des solutions envisagées... et inévitablement des problématiques. Les questions au centre de toute tentative de théorisation demeurent, elles, les mêmes : quelles unités linguistiques devons-nous prendre en compte pour garantir une segmentation de l'oral ou de l'écrit en unités homogènes ? Faut-il se passer de certaines catégories, et si oui, lesquelles et pourquoi ? À quel niveau opèrent les définitions et critères retenus ? Enfin, la phrase reste l'objet de discord principal entre linguistes, sa pertinence ayant été mise en doute (pour l'oral) essentiellement depuis les travaux du Groupe aixois (GARS).
- 2 L'ouvrage dirigé par Béguelin, Corminboeuf et Lefeuvre n'a pour ambition ni de proposer des solutions définitives ni de fournir une synthèse de la diversité des points de vue ; ambition qui, selon les éditeurs, tous les trois parfaitement au fait des problèmes qui sous-tendent les procédures de segmentation, « semble pour le moment hors de portée » (p. 10). L'intérêt du volume réside ailleurs : les contributions réunies couvrent quasiment tous les domaines dans lesquels les unités de segmentation en vigueur sont convoquées. Ainsi, le lecteur aura un panorama complet des analyses syntaxiques, sémantiques, discursives, textuelles, didactiques, psycholinguistiques, outillées, dont les entités qui servent à la segmentation font l'objet. Le tout est bien

pensé et servi par une organisation du livre en trois parties et onze chapitres parfaitement équilibrés. La divergence entre les modèles y est assumée et l'émergence de développements ultérieurs y est signalée. Le travail éditorial est à saluer : un avant-propos (pp. 9-12) concis et éclairant, une bibliographie conséquente réunie en fin de volume (pp. 221-238) suivie d'un *index nominum* (pp. 239-243) et d'un *index rerum* (pp. 245-255) toujours très appréciés par les lecteurs. Enfin, nous ne relevons aucun faux pas orthotypographique.

- 3 Pierre le Goffic ouvre la première partie du volume (« Enjeux épistémologiques »). Son article intitulé *De la langue au texte : à la recherche des unités perdues* (pp.15-32) expose avec une pédagogie magistrale les tenants et aboutissants des problèmes qui se rapportent à la (ré)conciliation du trio langue/discours/texte mis en regard avec trois phases du langage : le langage en puissance, son activité et son résultat. L'auteur montre comment les unités relevant de niveaux différents (de langue vs du texte) ont vu leur contenu renforcé par la reprise d'éléments appartenant à la tradition sans être néanmoins repensés (cf. aussi à ce propos Melis 2000). Son angle d'attaque sous forme de question (« peut-on reconnaître les mêmes unités dans les trois ordres de réalités¹ ? » ; p. 17) pose clairement l'hétérogénéité des unités impliquées dans la segmentation de textes. Le développement révèle pas à pas au lecteur les biais qu'engendre l'application de « procédures d'analyse statiques à un objet dynamique » (p. 20) lorsqu'on ne tient pas compte de la dimension temporelle inhérente à tout discours. Cet écueil conduit au brouillage des repères discursifs du propos des locuteurs, lequel doit concilier « le développement linéaire » et « la structuration hiérarchique ». Le Goffic fait une incursion dans la tradition grammaticale et rhétorique jusqu'à l'avènement du texte et aux besoins nouveaux liés aux dimensions du traitement de grands corpus. Il procède ensuite à la critique des différents modèles en vigueur et montre qu'on se trouve face à un changement de « paradigme scientifique ». Il n'est pas question pour lui de dire « adieu à la phrase »² ; au contraire, il lui laisse sa place d'unité « robuste », « simple » et « la plus satisfaisante » (p. 29). Il ne reste plus qu'à « élaborer un véritable modèle du discours » (p. 31) qui tienne compte des difficultés liées au traçage des limites entre langue et discours. Cet article nous permet de prendre du recul par rapport à ce qui est enseigné d'ordinaire de manière assez mécanique et nous aide à repenser les notions de *langue*, *discours* et *texte*, que nous considérons, à tort, aller de soi.
- 4 Dominique Legallois examine aussi les rapports entre grammaire et discours mais cette fois par le prisme de la phraséologie, des routines formulaires et de la structuration textuelle. Son article *Les constructions grammaticales comme schémas pré-discursifs* (pp. 33-44) s'inscrit certes, dans les grandes lignes, dans la mouvance des Grammaires de Constructions mais s'appuie également sur les travaux de linguistique fonctionnelle de M.A.K. Halliday tout en faisant appel à des notions issues de la théorie gestaltiste ou encore de l'analyse conversationnelle et ethnométhodologique (cf. les « first verbs » de Sacks 1992 ; p. 42). On reconnaît bien ici l'éclectisme de l'auteur qui parvient à lier élégamment différentes notions qu'il fait converger vers un seul objectif : exemplifier la systématité de la (pré)disposition de certaines unités en langue à être et/ou à devenir « des préfigurations du discours » (p. 44). Legallois montre le rôle des motifs (au sens narratologique) dans l'émergence « des constructions dont la fonctionnalité discursive est essentiellement d'établir un certain type de relations portant sur les participants du discours » (p. 32). Son dessein n'est pas de trancher en faveur de telle ou telle unité qui serait constitutive de la segmentation des textes, mais de montrer la

nécessité de considérer que les schémas narratifs (par exemple des séquences préparatoires ou de clôture) et les schémas actanciels sont des unités mémorisées par les locuteurs et qu'ils font partie de leurs compétences grammaticales. Nous retenons, comme c'était aussi le cas dans le premier article du volume, que vouloir à tout prix dissocier la compétence grammaticale des locuteurs de leur compétence discursive est un faux problème : les procédés constructionnels sont des objets de « connaissance préétablie » formant un « savoir-faire préalable » (p. 43) et, en tant que tels, ils endossent une fonction interactionnelle, c'est-à-dire « un rapport entre locuteurs ».

- 5 La contribution de Caroline Masseron (pp. 45-63) souligne la nécessité d'articuler micro- et macro-syntaxe et porte sur les enjeux didactiques du repérage, de l'analyse et de l'apprentissage « des phénomènes d'encodage syntaxique récurrents » (p. 45). En effet, en s'appuyant sur un échantillon d'écrits scolaires, l'auteur met en lumière la difficulté des enseignants (et des apprenants) à dépasser le cadre de la phrase et à faire appel à une segmentation « en trois temps » (cf. en séquences narratives, en clauses et en périodes), qui serait bénéfique pour l'apprentissage des processus de cohésion et de progression textuelle, du rôle des temps verbaux ou encore des liens référentiels. Masseron attire l'attention sur le fait que ces phénomènes sollicitent des compétences solides en matière de méthode et sont tributaires de la mise en place d'une progression des activités proposées par l'enseignant. Comme dans l'ensemble des contributions de cette première partie, le propos souligne l'importance de ne pas perdre de vue que langue et discours vont de pair.
- 6 Le dernier article de la première partie, celui d'Annie Kuyumkuyan (*Récurtivité des unités dans les discours : enjeux épistémologiques et sémantico-pragmatiques* ; pp. 65-82), est également une illustration de la richesse des modèles convoqués dans le volume. L'auteur recourt à l'approche genevoise de Roulet et de son équipe et met en évidence la complexité des plans d'organisation du discours et des unités dont on dispose pour l'analyse linguistique. En s'appuyant sur l'examen des dialogues fictionnels, elle prône une observation fine des discours préalable à toute formulation d'hypothèses quant à la validité de telle ou telle unité de segmentation. Elle montre qu'en recourant aux notions de *thème suspendu*, de *sujet en quête de prédicat*, d'*acte illocutoire*, de *tour de parole*, de *clause* et évidemment celles d'« intervention » et d'« acte discursif principal et secondaire », autrement dit en « dialogisant la syntaxe » (p. 74), on est conduit à dépasser le clivage entre niveau pragmatique, communicatif et niveau formel, syntaxique. Kuyumkuyan prouve que procéder à des lectures multiples d'un même dialogue « repose sur une parenthésisation spécifique des termes du discours » (p. 80), sur la possibilité pour un échange d'« intégrer une intervention à titre de constituant » (p. 81), de telle sorte que la construction du sens soit appariée à la hiérarchisation des unités et aux plans d'information discursive. En amorçant l'analyse de phrases dites *segmentées*, cet article sert de passerelle à la suite du volume.
- 7 La deuxième partie de l'ouvrage (*Modèles d'analyse et choix de segmentation*, aux pp. 85-162) se concentre sur l'étude de phénomènes plus précis ou ponctuels qui mettent à mal la conception canonique de la phrase, tels que les clauses nominales, les greffes, les segments flottants, les unités résomptives/prédicatives et l'amorçage.
- 8 Alain Berrendonner (pp. 86-98) prend à contrepied la conception classique de la *prédication* et de la définition sémantique de la phrase en termes d'association d'un sujet à un prédicat à laquelle on attribue une modalité énonciative. Après avoir rappelé les origines et l'évolution de cette conception depuis Aristote, l'auteur expose, puis

discute méthodiquement, les arguments classiques qui contribuent à caractériser les clauses nominales « comme des formes phrastiques défectives » (p. 87). L'idée maîtresse de cet article est d'« inverser la perspective usuelle » (p. 98) qui consiste à ramener les exceptions à la règle : Berrendonner généralise le modèle nominal à tout type d'énoncé et montre que les clauses nominales et les clauses verbales obéissent aux mêmes contraintes interprétatives et partagent les mêmes caractéristiques. Elles servent à accomplir « un acte de monstration », à « montrer le nom d'un objet-de-discours » (p. 94), à « incrémenter la mémoire discursive d'un fait nouveau » ou à « réactiver un fait déjà connu » (p. 96). Cette contribution met également en garde contre les dérives dues aux amalgames entre effets pragmatiques et traits de sens invariants et plaide pour un retour à une « sémantique pauvre », dans le bon sens du terme, à savoir une sémantique capable de se défaire « de la tendance à verser dans le sens littéral des énoncés n'importe quel effet interprétatif observable en discours » (p. 98). Ce faisant, l'auteur réussit son pari en nous offrant un modèle qui illustre que tout énoncé, qu'il soit nominal ou verbal, « se réduit littéralement à un simple nom désignant un objet concevable » (p. 98).

- 9 Le chapitre suivant s'attaque aux problèmes de segmentation liés aux « greffes » (cf. par exemple [[j'ai fini] [j'avais à peu près dix-huit ans]]; p. 100) et aux segments « flottants » (cf. le constituant B en italiques : Un ami publicitaire, chez qui [je prenais]_A [un pot]_B [ça va, trois pots bonjour les dégâts]_C, s'étonnait...; p. 101). Marie-José Béguelin & Gilles Corminboeuf (pp. 99-129) procèdent au tri systématique des structures qui relèvent du phénomène général de l'ellipse syntagmatique, répertorient et discutent les cas problématiques en comparant les solutions apportées par le modèle aixois au modèle fribourgeois afin d'offrir aux lecteurs des critères pertinents pour le décompte des unités syntaxiques maximales et des actes énonciatifs. Ils examinent tour à tour les limites de la rection, la superposition des phénomènes de parenthésisation, les processus mis en œuvre lorsque le locuteur produit dans son discours des bribes, des hésitations, des ruptures, des abandons ou des reprogrammations, l'apport des notions de « mémoire discursive » et de « routines périodiques ». Les remarques syntaxiques sont couplées à des analyses prosodiques qui permettent d'établir le statut d'une construction. La prise en compte dans la description des cas dits « métanalytiques » est cruciale, car elle évite de « trancher artificiellement, en imposant une solution au détriment d'une autre » (p. 117). Elle permet en revanche de mieux codifier le statut des cas sous-spécifiés afin de disposer de critères fiables dans les pratiques de segmentation (cf. la possibilité de recourir à des multi-transcriptions, par exemple). Il s'agit d'un fait duquel tout chercheur est sûrement coutumier : lorsque pour un même phénomène deux analyses sont envisageables et qu'on ne dispose pas d'indices fiables pour lever l'ambiguïté, on est souvent amené à trancher en fonction du plan théorique ou du domaine linguistique envisagé. De toute évidence, le choix n'est pas sans conséquences pour l'analyse des faits étudiés. Or, il n'est que rarement explicité dans les travaux. Béguelin & Corminboeuf montrent justement les conséquences du choix de laisser dans l'ombre ce type d'information lorsqu'on a l'ambition de fournir des typologies fondées sur des cas pertinents. Les auteurs insistent enfin sur la nécessité de se débarrasser de l'idée reçue selon laquelle les « greffes » et les « segments flottants » seraient propres à l'oral.
- 10 À l'instar des contributions précédentes, l'article de Florence Lefeuve (pp. 131-147) est à l'interface de la syntaxe et du discours. Elle y examine les caractéristiques syntactico-

sémantiques et les valeurs discursives des constructions faisant appel à la configuration « unité résomptive/unité prédicative » autour de *c'est vrai* (cf. *c'est vrai, je t'ai un peu oublié* et ses variantes en *P*, *c'est vrai* ou *c'est vrai* intégré en *P*). L'auteur procède à une analyse détaillée de cette construction en étudiant minutieusement son environnement distributionnel et discursif ou ses traits sémantiques et en étayant sa description grâce à des tests syntaxiques systématiques. Elle montre que la combinaison de deux unités, résomptive et prédicative, forme une « unité particulière en discours » appelée « période discursive » possédant des valeurs multiples sur le plan argumentatif (cf. « renforcement, justification, exemplification, restriction », p. 147).

- 11 L'article de Georgeta Cislaru & Thierry Olive (pp. 149-162) clôt la deuxième partie du volume en offrant un éclairage inédit sur les phénomènes d'écriture dynamique appréhendés d'un point de vue processuel, en tant que « jets textuels » (i.e. « des unités linguistiques produites entre deux pauses lors du processus de rédaction » ; p. 149). Prenant appui sur des phénomènes d'amorçage, les auteurs interrogent le pouvoir projectif de *jets textuels* et expliquent comment ces derniers deviennent des jets de « complétion ». Cislaru et Olive définissent les amorces à la fois d'un point de vue linguistique et psycholinguistique. Leurs données sont issues de rapports éducatifs et académiques enregistrés en temps réel et analysés grâce aux logiciels Inputlog et Scriptlog qui permettent le suivi des frappes et la mesure des pauses révélatrices des « unités de performance écrite » (p. 154). Ils nous fournissent ainsi une typologie d'amorces (morphosyntaxiques et lexico-sémantiques) en se fondant sur des contraintes mémorielles et combinatoires (cf. l'étude de bornes droites : prépositions, déterminant, noms). Cet article ouvre plusieurs perspectives sur la compréhension et la modélisation de relations de dépendance en lien avec les performances langagières. Par son côté empirique, il amorce aussi la troisième et dernière partie de l'ouvrage : *Analyses sur corpus et/ou segmentation outillée*.
- 12 Le chapitre 9 est le seul du volume à s'intéresser à une autre langue que le français, en l'occurrence l'espagnol, et présente de ce fait un intérêt certain. Dans *Les fragments comme unités linguistiques : une analyse de corpus de l'espagnol oral* (pp. 165-180), Oscar Garcia Marchena, spécialiste de la question, procède à la distinction de deux types de structures sans verbe, les phrases averbales et les fragments. Les premières sont définies par la réalisation d'une tête autre que verbale (cf. par exemple, *A mi derecha, el ministro, José Luis Corcuera* ; p. 168), les secondes sont des formes d'ellipse (la tête prédicative et/ou des arguments manquent). Tout en admettant des difficultés de classification, l'auteur tente une catégorisation d'abord des phrases averbales sans ellipse (cf. existentielles, polaires et prédicatives), puis des fragments selon la « présence ou absence de corrélat » (p. 175) au niveau syntaxique et selon les types d'acte de parole, assertant ou questionnant, qu'ils réalisent. Nous regrettons l'absence d'une discussion ou d'un rappel des données à la base de la notion d'*ellipse* ou encore l'absence de renvoi à certains travaux comme ceux du volume dirigé par Hadermann *et al.* (2013), par exemple. Garcia Marchena plaide pour l'utilité de la notion de *fragment*, à différencier de celle de *phrase elliptique*. Malgré un effort de systématisation des exemples, les fragments soumis à l'étude nous paraissent regrouper des phénomènes hétérogènes (cf. aussi Deulofeu 2013). De même, l'exclusion de la catégorie « fragments » de certaines structures « tronquées » (p. 168) comme les *fillers* ou *amorces* réparties sur deux tours de parole peut être discutée³. S'agit-il d'ellipses structurales ? D'incomplétudes de réalisation d'un énoncé ? L'auteur arrive toutefois à démontrer l'utilité d'affiner certaines propriétés syntactico-sémantiques qui semblent, en effet,

être propres aux séquences appelées *fragments* les distinguant ainsi des phrases verbales.

- 13 Les deux derniers articles du volume sont très éclairants à la fois en ce qui concerne les applications en traitement automatique des langues et le choix de tel ou tel modèle théorique à la base des annotations de vastes corpus. La segmentation précédant plusieurs tâches en TAL, elle constitue un défi pour l'apprentissage automatique.
- 14 Mathilde Carnol et Anne Catherine Simon (pp. 181-201), se fondant sur l'approche du groupe aixois de recherches en syntaxe, décrivent qualitativement et quantitativement, de manière minutieuse, les différentes formes de « dispositifs » de la rection (cf. les travaux de C. Blanche-Benveniste). Fortes de la longue expérience en la matière du centre de recherche VALIBEL, elles adoptent une méthode de segmentation syntaxique qui peut être menée indépendamment de la prosodie. Elles définissent dans une première partie toutes les notions utilisées pour l'analyse en « dispositifs » aux niveaux micro- et macrosyntaxique. La méthodologie et la présentation du corpus Locas-F (Louvain Corpus of Annotated Speech-French) mettent en exergue la nécessité de prendre en compte la variété des genres de discours afin de mieux appréhender « les différentes manières de disposer les éléments de la rection, selon les besoins discursifs » (p. 184). Les troisième et quatrième parties de l'étude s'attachent non seulement à fournir un inventaire des variantes de dispositifs de la rection verbale, mais aussi à les problématiser en montrant les limites de certains rattachements et en spécifiant les types de constituants syntaxiques concernés par chaque dispositif. Cet article montre enfin, grâce au caractère exhaustif de l'annotation suivie, que « l'ordre des mots sujet-verbe-complément est une norme objective de la syntaxe du français parlé » (p. 201).
- 15 Dans la même perspective, Nathalie Rossi-Gensane, Biagio Ursi, Iris Eshkol-Taravella et Marie Skrovec (avec la collaboration de Margot Lambert et de Luisa Acosta Cordoba ; pp. 203-220) exposent les enjeux de la segmentation (« en empirie et en théorie ») en comparant trois projets (Rhapsodie, Orféo et SegCor) et les choix de segmentation adoptés afin de disposer de protocoles d'annotation satisfaisants. Les auteurs détaillent d'abord les notions de « rection », « noyau », « adnoyau », etc. issues du cadre théorique aixois, puis ils présentent une autre unité d'analyse (employée dans le projet SegCor), celle du *chunk/chunking*. Le « chunk » est au départ une notion psycholinguistique et renvoie à une unité d'information qui intervient en mémoire à court terme (cf. regroupement de l'information en « paquets » ; Miller 1956). Dans son article de 1991, Abney applique ce concept à la segmentation ('parsing') en postulant que « lorsqu'on lit une phrase, on la lit morceau par morceau »⁴. Il s'agit d'une unité d'information constituée de segments non récursifs dont le prototype comprend un mot plein entouré par des mots fonctionnels vérifiant un motif (cf. « phrases » adjectivales, adverbiales, nominales, prépositionnelles, verbales ou encore des unités illocutoires autonomes comme *voilà*, *d'accord*, etc. ; p. 212). Un autre point fort de cet article est de souligner qu'en fonction des critères choisis et du statut qu'on attribue à une unité (cf. micro- ou macro-syntaxique), l'analyse qui en découle est fortement différente d'un projet à un autre ; de ce fait, une systématisation des données doit être effectuée. Rossi-Gensane *et al.* fournissent à ce propos des solutions et des propositions « en vue de la finalisation de protocoles d'annotation » (p. 220).
- 16 Dans l'ensemble, *Types d'unités et procédures de segmentation* contient des études pointues qui nécessitent toutefois de la part du lecteur quelques connaissances préalables sur le

sujet traité. Il s'agit d'un livre d'une grande qualité qui parcourt tout le champ d'investigation du domaine de la segmentation et qui introduit les références nécessaires à l'appréhension de tous les phénomènes évoqués ici. Les éditeurs de cet ouvrage ont pris leurs précautions en indiquant ne pas pouvoir fournir une synthèse des travaux portant sur les choix des unités pour la segmentation de textes. Ceci n'est toutefois que partiellement exact, car comme on a pu l'apercevoir dans ce compte-rendu, le lecteur trouvera non seulement les dernières avancées dans le domaine, mais aussi un « portrait » des unités d'analyse et une esquisse « des difficultés auxquelles elles achoppent » (quatrième de couverture). Il s'agit indéniablement d'un livre, qui grâce à la confrontation des étiquettes et des modèles, pose des bases solides pour l'acquisition et/ou la poursuite des observations théoriques et pratiques.

BIBLIOGRAPHIE

Abney, S. (1991), "Parsing by Chunks", in R.C. Berwick, S. Adney and C. Tenny (eds.), *Principle-Based Parsing*, Dordrecht, Kluwer Academic Publisher, 257-278.

Deulofeu, J. (2013), « De l'inutilité de la notion de 'fragment' pour la description des énoncés 'fragmentés' », in P. Hadermann et al. (dir.), *Ellipse & fragment. Morceaux choisis*, Berne, Peter Lang, 157-180.

Hadermann, P. Pierrard, M. Roig, A. & Van Raemdonck, D. (dir.) (2013), *Ellipse & fragment. Morceaux choisis*, Berne, Peter Lang.

Kleiber, G. (2003), « Faut-il dire adieu à la phrase ? », *L'information grammaticale* 98, 17-22.

Melis, L. (2000), « La phrase et son analyse », *Modèles linguistiques* 42, 79-145.

Miller, G.A. (1956), "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information", *Psychological Review* 63, 81-97.

NOTES

1. Les trois ordres de réalité sont : une réalité neurologique (la langue), une activité individuelle et sociale (le discours) et un produit matériel (le texte).
2. Kleiber (2003).
3. Nous trouvons ici une position différente de celles défendues dans les études de Kuyumkuyan et de Cislaru & Olive.
4. "I begin with an intuition: when I read a sentence, I read it a chunk at a time. For example, the previous sentence breaks up something like this: [I begin] [with an intuition]: [when I read] [a sentence], [I read it] [a chunk] [at a time]" (Abney 1990: 257).

AUTEUR

HÉLÈNE VASSILIADOU

Université de Strasbourg