



HAL
open science

De la langue au texte : à la recherche des unités perdues

Pierre Le Goffic

► **To cite this version:**

Pierre Le Goffic. De la langue au texte : à la recherche des unités perdues. Types d'unités et procédures de segmentation, Lambert-Lucas, pp.15-32, 2020, 9782359352870. hal-03941355

HAL Id: hal-03941355

<https://hal.science/hal-03941355>

Submitted on 16 Jan 2023

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.

De la langue au texte : à la recherche des unités perdues P. Le Goffic

Au départ, une interrogation née du titre du colloque : quel rapport y a-t-il entre les 'types d'unités' linguistiques et les procédures de segmentation ? Qu'entend-on au juste par 'unités' linguistiques ? De quelque façon qu'on les prenne, les unités linguistiques sont-elles *ipso facto* des unités de segmentation du texte ? Y a-t-il des unités linguistiques propres à segmenter le texte ?

Pour essayer d'y voir clair, on procédera en trois temps. Dans une première partie, on posera au plan théorique la question des unités, en relation avec la tripartition langue / discours / texte. Dans la deuxième partie, historique, on verra comment la grammaire a apporté des réponses différentes à cette question selon les époques, et en fonction de son rapport au texte. On essaiera enfin de tirer les leçons de cette enquête pour aborder du mieux qu'il est possible la question des unités, - de la langue comme du texte -, à l'heure de la 'linguistique de corpus'.

1. Unités de langue vs unités du texte

La problématique de départ semble aussi naturelle qu'incontournable : « Voici un texte. Découpons-le en unités ». Comme en témoigne ce colloque lui-même, la segmentation des textes (poussée aux dimensions du traitement de corpus considérables) est devenue une activité fondamentale, au cœur de la pratique des grammairiens et des linguistes contemporains.

Or il est frappant de constater que c'est là une problématique récente dans l'activité et les préoccupations des grammairiens : on peut dire qu'elle date de moins de deux siècles, durée qui est certes loin d'être négligeable, mais représente peu de choses si on la rapporte à l'histoire des recherches sur le langage, inaugurée par les Grecs, qui excède largement deux millénaires. Pour rester plus près de nous, il est frappant de constater que la question ne se posait nullement pour les classiques : aucun grammairien de l'époque classique n'a jamais songé à découper un texte quel qu'il soit (littéraire ou non) comme nous le faisons. Pourquoi ? Manque général d'intérêt pour les textes ? Certes non : on sait combien certains genres au moins étaient cultivés, et leurs productions étudiées, commentées, analysées, - mais jamais « segmentées ».

Que s'est-il passé ? On est typiquement dans le cas d'un changement de paradigme scientifique : ce qui change, ce ne sont pas tant les réponses aux questions qu'on se pose, que les questions elles-mêmes, les fondements sur lesquels elles s'articulent et se légitiment. Quels sont ces fondements ? L'articulation entre la langue, le discours et le texte. Ces trois notions sont omniprésentes dans la réflexion sur le langage depuis ses débuts, avec un statut ambivalent, apparaissant tantôt comme triviales (banalités qu'il ne vaut même pas la peine d'énoncer), tantôt comme problématiques, source de difficultés embarrassantes voire insolubles, quelque peu suspectes de métaphysique et dont on peut faire l'économie. Dans les deux cas, mieux vaut passer rapidement ...

Là est pourtant le cœur du problème. Et pour aborder les relations entre la langue, le discours et le texte, il faut remonter à celui des rapports entre la **puissance** et l'**acte**. L'opposition, on le sait, remonte à Aristote, qui cherche à en démontrer le bien-fondé et à en voir les manifestations du côté du mouvement et de la physique, sans jamais évoquer le langage à ce sujet (- ce que les linguistes ne peuvent que regretter amèrement !). Mais Humboldt, et d'autres à sa suite, dont Chomsky, en ont repris l'idée et l'ont étendue au

langage, qui l'illustre de façon typique : la distinction entre la puissance et l'acte s'applique pleinement au langage : elle conduit à distinguer langue, discours et texte. Dans cette ligne, on posera ici quelques définitions de base, minimales (sur lesquelles il est permis d'espérer que le consensus soit large sinon unanime)

Le **langage en puissance**, c'est le fait de pouvoir parler, la **faculté de langage**, telle qu'elle existe d'abord dans la tête du bébé qui vient de naître, et qu'elle devient ensuite, après apprentissage, la faculté de parler et de comprendre une langue (ou plusieurs). Le terme chomskyen de compétence convient bien, évoquant un mécanisme qui permet de parler. On peut aussi utiliser simplement, comme nous le faisons ici, le terme de **langue**, au sens de système de règles intériorisé et partagé par les locuteurs d'une communauté linguistique. La langue (ou compétence) ainsi entendue est une réalité abstraite qui siège dans le cerveau, et qui est non observable (du moins en l'état actuel des procédés d'investigation, et pour longtemps sans doute). Les grammaires et les dictionnaires que produisent les linguistes sont des tentatives d'objectivation, d'extériorisation, de cette connaissance.

Le **langage en acte** se présente sous deux aspects : l'activité en elle-même et le produit qu'elle engendre.

L'activité, c'est le **discours** (ou performance), c'est-à-dire ce que font les sujets parlants : parler (ou écrire), écouter (ou lire), communiquer entre eux, en temps réel, par le moyen du langage.

Le produit, c'est le **texte** : résultat extériorisé, matériel de l'activité de langage. Toute activité de langage visant à la communication se traduit, par définition, par un texte, qui est une réalité physique, aussi bien dans le cas de l'oral que de l'écrit. Dans le cas de l'oral, le texte, c'est le flux des ondes sonores qui part du locuteur et que capte le récepteur ; ce texte est par essence volatile et fugace, - à moins qu'un moyen quelconque ne permette de le fixer et de le conserver : on y reviendra.

Le rapport entre ces trois termes est très particulier : ils sont inséparablement liés (le langage suppose les trois) tout en étant extraordinairement hétérogènes : comment traiter à l'identique une réalité neurologique dans la tête (la langue), une activité à la fois individuelle et sociale, avec mille déterminations (le discours), et un produit matériel (texte : ondes sonores, traces sur papier, ...) ? Ces différences doivent nous interpeller : peut-on reconnaître les mêmes unités dans les trois ordres de réalités ? Si on prend au sérieux leur hétérogénéité, il ne va pas de soi qu'on puisse trouver un dénominateur commun, une mesure commune, à la langue, au discours et au texte.

Il convient de s'arrêter spécialement sur le cas du texte, qui est le seul des trois composants du langage à être physiquement observable, mais qui est aussi celui dont le statut est le moins clair. Sa matérialité même a fait du texte un produit historiquement variable : pendant 95% au moins de l'histoire de l'*homo sapiens* moderne (qui s'identifie à l'*homo loquens*), le texte a été purement oral, réalité sonore fugace, perceptible dans le seul instant de sa production et ne laissant pas de traces physiques au-delà du moment même de la profération. Avec l'invention de l'écriture (longtemps après l'apparition du langage, mais il y a peu par rapport à nous) et sa diffusion, un autre type de texte est apparu, qui – fait crucial – se détache du discours qui l'a fait naître, et se présente sous une forme physique radicalement différente. Et l'oral peut maintenant être fixé, rendant caduc le vieux dicton selon lequel *verba volant*, les paroles s'envolent. Au total, si le statut de la langue et celui du discours n'ont pas évolué depuis les débuts du langage, le texte apparaît dans ses différentes formes comme un produit historiquement variable, dont le statut est loin d'être évident, peut être sujet à variations, et n'a sans doute pas retenu toute l'attention qu'il mérite.

A ce stade, pour expliciter les rapports entre langue, discours et texte, nous allons considérer un échantillon de texte (oral transcrit, repris d'un exposé fait par Syvain Kahane à

Paris III), présenté ici d'une façon informelle qui n'a pour but que d'en faciliter la lecture. L'échantillon est assurément un texte, à partir duquel on peut aisément retrouver la distinction entre langue, discours et texte.

*nous essayons tant bien que mal de lui faire comprendre
que sa cabane
dans quelques années
entre parenthèses elle a 79 ans
quand elle aura des difficultés
ce qu'on (n') espère pas
des difficultés à se déplacer
ou à évoluer
c'est-à-dire qu'il y a énormément
d'escaliers à monter pour arriver à sa cabane*

donc

*le jour où elle ne pourra plus se déplacer
ou qu'elle sera malade un petit plus sévèrement*

*on essaye de lui faire comprendre
qu'elle ne pourra plus vivre dans cette cabane*

Discours. La lecture de cet extrait nous conduit spontanément à reconstituer d'une certaine façon le déroulement d'un discours dans le temps.

Le discours commence d'une façon extrêmement canonique, avec des structures complexes ouvrant sur une subordonnée complétive, dont la charnière sujet – prédicat reste en suspens (*nous essayons tant bien que mal de lui faire comprendre que sa cabane dans quelques années ...*). Le circonstant *dans quelques années* déclenche très vraisemblablement l'incise explicative *entre parenthèses elle a 79 ans*, et oriente vers les difficultés liées à l'âge (*quand elle aura des difficultés ...*). L'exposition des *difficultés* est coupée par une autre incise (dans laquelle le locuteur se met en scène, avec le souci de soigner son image : *ce qu'on (n') espère pas*), puis reprise (*difficultés à se déplacer ou à évoluer*), en appelant au passage une longue explication (*c'est-à-dire qu'il y a énormément d'escaliers...*).

Alors que le programme inauguré par le *que* complétif (*on essaye de lui faire comprendre que ...*) n'a toujours pas été mené à son terme, et que différents topiques secondaires (l'âge de l'occupante, les escaliers, ...) ont été touchés, le mot *donc* est sans doute la marque que le locuteur ressent la nécessité de reprendre en main la conduite de son discours, comme une sorte de 'remise à zéro' en vue d'une restructuration. La restructuration n'est que de faible amplitude : la série des circonstants de temps reprend (*le jour où elle ne pourra plus ...*) et, au moment d'arriver enfin au message final de la complétive, le locuteur, par précaution, reprend presque *verbatim* son introduction initiale (*on essaye de lui faire comprendre que ...*), pour livrer enfin son message final (*elle ne pourra plus vivre dans cette cabane*).

Cet extrait illustre bien ce que c'est que le discours ou activité de langage : il nous montre les efforts du locuteur aux prises avec sa propre intention de signification (perpétuellement retouchée, complétée, infléchie, ...), cherchant à la couler dans une forme linguistique contraignante et perpétuellement soumise à auto-évaluation, luttant contre le temps dans une sorte de course contre la montre : il a besoin du temps pour développer et/ou corriger son propos, mais le temps éloigne des repères structurants initiaux et risque de les oblitérer : le développement linéaire et la structuration hiérarchique sont des forces contraires qu'une énonciation réussie doit concilier.

Langue. Dans l'équipement cognitif considérable (capacité de traitement en parallèle, d'anticipation, de gestion des interactions, etc.) que nécessite une telle entreprise figure en bonne place la compétence linguistique, ce que nous avons appelé ci-dessus la 'langue'. Et notre extrait nous permet de nous faire une idée de la compétence linguistique du locuteur : il est manifeste qu'il connaît et maîtrise les structures du français (prédication, subordination, jeu sur les dislocations, etc.) et qu'il est soucieux de produire des énoncés conformes à un certain modèle. Si la poursuite de son projet de signification l'amène à sortir quelque peu des rails d'une énonciation totalement 'modèle', ses écarts sont toujours réalisés avec le souci de ne pas casser le fil syntaxique, et il gère ses libertés dans le respect d'un plan d'ensemble. Le seul point sur lequel il ne parvient pas à sauvegarder pleinement la consistance syntaxique de son énonciation concerne *sa cabane* : la répétition lexicale (en fin d'extrait) est peu compatible avec l'occurrence du mot en dislocation gauche dans la complétive (voir la deuxième ligne de l'extrait), mais la distance est grande, le mot *cabane* a déjà été 'sorti' de son cadre d'emploi initial et réutilisé à neuf dans une incise (*pour arriver à sa cabane*) et à ce point de l'énoncé il n'a plus un degré de saillance suffisant pour être anaphorisé simplement par un pronom (qui eût été un pari trop risqué sur la mémoire de l'interlocuteur).

Au total, le locuteur a une compétence linguistique solide : il a dans la tête un modèle de compétence bien installé (qui ressemble fort à celui de la phrase).

Texte. Reste à examiner pour lui-même le texte, qui est à la fois le plus évident des trois aspects du langage (le seul matériellement observable) et celui dont le statut est le plus incertain.

Ce texte est fait de traces écrites sur le papier, qui sont comme une suite de photos saisies au vol de ce que le locuteur a produit : la suite des mots qu'il a prononcés, l'un après l'autre, retranscrits *verbatim*, comme les 'minutes' de sa déposition. Cette fixation artificielle (aboutissant à un objet matériellement observable de conservation illimitée) n'est pas constitutive du langage.

Ce texte, en l'état, est-il le reflet exact et fidèle de l'intention de signification du locuteur ? Il est clair que non : l'intention de signification du locuteur ne saurait être identifiée strictement à la suite des mots qui ont été proférés, avec ses interruptions, ses redites, ses retours en arrière, etc. : on ne saurait soutenir que le locuteur a 'voulu' se répéter ou faire des ruptures de construction, etc.

De l'autre côté, ce texte correspond-il mot pour mot, rigoureusement, à ce que l'interprète comprend ? Non, pour les mêmes raisons : il n'y a nul doute que l'interlocuteur réinterprète activement ce qu'il entend, en 'gommant' les aspérités, et se forge une représentation de l'intention de signification de l'émetteur, reconstruite (virtuellement reformulée) à partir de ce qu'il a entendu.

Mais alors qu'est-ce au juste que ce texte, si ce n'est ni ce que le locuteur a voulu dire, ni ce que l'interlocuteur a compris ? Ce ne peut être que la trace de l'effort du locuteur pour mettre en mot son intention de signification en temps réel, une médiation destinée à l'auditeur pour que celui-ci s'en saisisse et (re)construise son interprétation.

Il se trouve que la technologie qui est présentement la nôtre permet que cette trace fugace soit fixée, captée, conservée, qu'elle soit dotée artificiellement d'un visage stabilisé, où sont co-présents des éléments qui n'étaient pas destinés à l'être. Cette présentation annule la dimension du temps (primordiale dans l'usage naturel du langage) et risque de nous faire appliquer des procédures d'analyse statiques à un objet dynamique, comme si nous voulions décrire en deux dimensions un objet tridimensionnel.

Cette annulation de la dimension du temps par un texte arrêté, fixé, n'est pas valable uniquement pour la transcription d'un texte oral : elle vaut inéluctablement pour tout texte, y compris les textes explicitement destinés à circuler sous forme écrite (littérature, presse, ...).

Unités. Qu'en est-il alors des unités ? La question primordiale est : des unités de quoi ? de la langue, du discours, ou du texte ? Impossible en effet de faire l'économie de cette question, qui doit être posée, - alors qu'elle ne l'est pas toujours, et que sa formulation même est délicate.

2. Variation historiques.

L'histoire peut ici fournir des éléments de réflexion pertinents. Nous distinguerons deux paradigmes successifs :

1. le paradigme classique : la langue (dans la dépendance de la pensée) repose sur la proposition (expression d'un jugement), et le discours n'est que la mise en œuvre d'une série de propositions (éventuellement altérées par les 'figures'). Le texte n'a pas d'existence distincte du discours. Les seules unités linguistiques sont donc celles de la langue, à savoir la proposition.

2. le paradigme moderne (depuis le 19^e siècle) : par-delà ses formes diverses, il se caractérise globalement par la question de l'émergence du texte en tant qu'instance du langage et objet d'étude pour lui-même. Cette émergence s'est produite en deux temps : d'abord par la promotion du texte écrit au 19^e concomitante avec l'installation de la phrase comme concept organisateur général, unité à la fois de la langue et du texte, puis, dans la deuxième moitié du 20^e, par l'irruption du texte oral, entraînant la remise en question de la phrase (à défaut de retour sur la problématique). Les développements contemporains de la 'linguistique de corpus' ne font que confirmer et accentuer la place et l'importance du texte.

Face à ces évolutions, les linguistes, depuis le début du 20^e siècle, ont multiplié les controverses, et on tentera d'en dire un mot, avant de risquer quelques observations à visée prospective.

2.1. Le paradigme classique : de la langue au discours, la proposition

Pour les classiques, les unités pertinentes sont les unités de la langue (qui est liée à la pensée, et ne se distingue pas du langage) ; le discours n'en est que la mise en œuvre et son étude consiste principalement à y retrouver les unités de la langue. Le discours relève donc à la fois de la grammaire, pour ce qui est d'en dégager les propositions dont il est constitué, de la logique, pour ce qui concerne la combinaison des diverses propositions, et de la rhétorique, pour ce qui concerne les écarts par lesquels les hommes s'écartent de la formulation naturelle des propositions, et l'étude de certains assemblages harmonieux comme les périodes. Le texte n'a pas d'existence propre, distincte du discours, auquel il est complètement ramené et identifié.

On rappellera ici brièvement des choses bien connues, en prenant appui sur la *Grammaire* et la *Logique* de Port-Royal : Arnauld et Lancelot, *Grammaire générale et raisonnée*, 1660, et Arnauld et Nicole, *La Logique ou l'Art de penser*, 1662.

Langue. Les logiciens et grammairiens de Port-Royal ne craignent pas d'affirmer la dépendance directe du langage par rapport à la pensée, et se font forts d'exposer les mécanismes de l'esprit qui nous permettent de parler :

« Que la connaissance de ce qui se passe dans notre esprit est nécessaire pour comprendre les fondements de la grammaire » (*Grammaire...*, Seconde partie, Titre du Chapitre premier, éd. 1969, p. 22).

Le mécanisme, tel que l'expose la *Logique*, est le suivant :

« Après avoir conçu les choses par nos idées, nous comparons ces idées ensemble, et trouvant que les unes conviennent entre elles et que les autres ne conviennent pas, nous les lions ou déliions, ce qui s'appelle *affirmer* ou *nier*, et généralement *juger*.

Ce jugement s'appelle aussi *proposition*, et il est aisé de voir qu'elle doit avoir deux termes : l'un, de qui l'on affirme, ou de qui l'on nie, lequel on appelle *sujet* ; et l'autre que l'on affirme ou que l'on nie, lequel s'appelle *attribut* ou *Praedicatum*.

Et il ne suffit pas de concevoir ces deux termes ; mais il faut que l'esprit les lie ou les sépare. [...] Ainsi quand je dis, *Dieu est juste*, Dieu est le sujet de cette proposition, et juste en est l'attribut, et le mot *est* marque l'action de mon esprit qui affirme, c'est-à-dire qui lie ensemble les deux idées de Dieu et de juste comme convenant l'une à l'autre » (*Logique...*, Seconde partie, Chap. iii, éd. 1993, p. 113).

On retrouve la proposition héritée d'Aristote, conçue comme répondant à une nécessité logique : c'est la forme linguistique du jugement, opération fondamentale de la pensée. Il existe certes des énonciations qui ne sont pas des jugements, mais le jugement est l'opération primordiale de la pensée, et il s'identifie totalement à la proposition : aucune autre forme d'expression du jugement n'est possible ni envisageable. Il n'y a donc pas lieu de chercher à falsifier ce postulat (comme nous dirions en termes modernes), en cherchant des contre-exemples ou des exceptions au modèle de la proposition.

Discours. Le discours est l'activité même de langage. Il repose nécessairement sur les mêmes bases, c'est-à-dire la proposition : parler, c'est énoncer des jugements (ou propositions). Comme le dit la *Rhétorique* de Lamy (publiée en 1675, et souvent considérée comme un prolongement de la *Grammaire* et de la *Logique* de Port-Royal) :

« Le discours n'est qu'un tissu de plusieurs propositions. » (Lamy, *Rhétorique*, I, X, éd. 1998, p. 89)

Dumarsais écrit encore en 1754 dans l'article 'Figure' de l'*Encyclopédie* :

« [...] il faut toujours réduire à la forme de la proposition toutes les phrases particulières et tous les membres d'une période. » (1988, p. 327 ; *phrase* a ici le sens classique et vague d'expression, et non le sens moderne)

L'étude du discours consiste donc à y retrouver (ou à y reconstruire si besoin est) des propositions, conformes à ce qu'elles doivent être, c'est-à-dire composées d'un sujet et d'un prédicat : telle est la tâche de la Grammaire. Et le champ de la Grammaire se trouve du même coup précisé et circonscrit : tout ce qui relève d'une combinaison de propositions (ainsi quand deux propositions sont reliées par une conjonction comme *et*, ou *si*, *parce que*) est du ressort de la Logique. D'autre part tout ce qui signale une distance, un écart entre le modèle propositionnel (considéré comme reflétant l'expression naturelle de nos pensées) et les productions effectives des hommes relève de la Rhétorique et de la théorie des Figures.

Voici comment la Grammaire de Port-Royal, dans ses dernières pages, concède, de mauvaise grâce, semble-t-il, une place aux Figures :

« Ce que nous avons dit ci-dessus de la syntaxe suffit pour en comprendre l'ordre naturel, lorsque [...] le discours] est conforme à l'expression naturelle de nos pensées.

Mais parce que les hommes suivent plus souvent le sens de leurs pensées, que les mots dont ils se servent pour les exprimer, et que souvent, pour abrégé, ils retranchent quelque chose du discours, ou bien que, regardant à la grâce, ils y laissent quelque mot qui semble superflu, ou qu'ils en renversent l'ordre naturel ; de là est venu qu'ils ont introduit quatre façons de parler, qu'on nomme *figurées*, et qui sont comme autant d'irrégularités dans la Grammaire, quoiqu'elles soient quelquefois des perfections et des beautés dans la langue ». (*Grammaire...* II, Chap. xxiv, Des Figures de Construction, éd. 1969, p. 106) :

Ces quatre figures, brièvement caractérisées, sont la SYLLEPSE, ou *Conception*, l'ELLIPSE, ou *Défaut*, le PLEONASME, ou *Abondance*, l'HYPERBATE, ou *Renversement*.

Au siècle suivant, Dumarsais a beau jeu de remarquer que les façons de parler par figures sont plutôt la règle que l'exception :

« On dit communément que les figures sont des manières de parler éloignées de celles qui sont naturelles et ordinaires [...]. [mais en réalité] il n'y a rien de si naturel, de si ordinaire et de si commun que les figures dans le langage des hommes. [...] Ainsi, bien loin que les figures s'éloignent du langage ordinaire des hommes, ce seraient au contraire les façons de parler sans figures, qui s'en éloigneraient, s'il était possible de faire un discours où il n'y eût que des expressions non figurées. (Dumarsais, *Des Tropes*, 1730, éd. 1988, p. 62 – 63).

Mais cette observation n'enlève rien au statut 'naturel' de la proposition.

Le discours est également considéré sous l'angle de son efficacité ou de sa réussite esthétique : il y a des arrangements particulièrement harmonieux, que cherche à caractériser la Rhétorique. C'est le cas en particulier de la '**période**', qui n'est pas une 'unité linguistique' à proprement parler et n'est pas dans un rapport précis par rapport à la proposition ; c'est une sorte d'unité subjective du discours : une période doit être constituée de plusieurs membres et pouvoir être énoncée d'une seule respiration.

Quant au **Texte**, il se fonde entièrement dans le discours (au point que 'discours' est souvent pris au sens de 'texte') : le résultat produit est complètement ramené à l'activité qui l'a produit (le discours) et ne constitue pas un objet d'étude en lui-même. Pour analyser les textes, la démarche classique est d'y rechercher ou d'en reconstruire les propositions qui le composent, en s'aidant au besoin des figures de la Rhétorique pour rendre compte de tout ce qui s'écarte de 'l'ordre naturel'. L'analyse peut être ensuite prolongée en direction de la Logique (pour la combinaison des propositions).

Unités : pour les classiques la question des unités linguistiques est donc simple et claire : les seules unités linguistiques sont les unités de la langue, c'est-à-dire (au-delà des mots) les propositions. Le discours (et le texte à sa suite) peut suivre à la lettre le modèle (qui s'identifie à l'ordre naturel du langage), ou s'en écarter superficiellement, par des figures (derrière lesquelles se retrouve le modèle).

2.2. L'avènement du texte

Pourquoi les choses ont-elles changé ? Parce que le texte s'est imposé dans le champ de la linguistique comme état de langage à prendre en compte pour lui-même, et qu'on a cessé de le voir comme un simple produit du discours.

Cette émergence s'est faite en deux temps : d'abord la promotion du texte écrit, au 19^{ème}, et ensuite l'irruption du texte oral, dans la deuxième moitié du 20^{ème}.

2.2.1. La promotion du texte écrit et l'émergence de la phrase (19^{ème} s.)

La promotion du texte au 19^o est due à des facteurs historiques et politiques, comme Chervel (1981) l'a bien montré : la pression de la démocratisation et de la généralisation de l'enseignement de la lecture et de l'écriture dans le pays. D'où la place du texte écrit, comme modèle exemplaire à proposer (il s'agit bien entendu des 'bons auteurs' de notre Panthéon littéraire, dans lesquels on trouve aisément des illustrations du français qu'on veut enseigner), et comme terrain d'exercices privilégié pour l'entraînement des élèves. Pour arriver au but visé, il fallait une grammaire claire, relayée par les instituteurs, eux qui ont la mission de transmettre la pratique correcte de la langue, et en particulier la maîtrise de l'orthographe, et acceptée par les élèves. L'heure n'est plus aux spéculations sur les rapports entre la langue et

la pensée, mais à la recherche d'un modèle pratique, auto-suffisant (sans recours à la logique), et conduisant directement à l'écriture d'un texte fait de phrases bien formées, écrit dans une orthographe irréprochable.

Par bonheur, le travail critique des grammairiens du 18^{ème} siècle (qu'on ne fera ici qu'évoquer d'un mot) fournissait les aménagements théoriques nécessaires : la notion de proposition était devenue moins rigide et uniforme, avec notamment l'élaboration de la notion de fonction grammaticale, l'inventaire des différents types de propositions, etc.

De la convergence de ces facteurs (externes et internes) est résultée la notion de phrase (au sens moderne), qui est devenue au fil du 19^{ème} le concept organisateur de toute la grammaire, et s'est imposée à la fois comme unité de langue et comme unité textuelle.

Langue. En tant qu'unité de langue, la phrase est une version assouplie et étendue de la proposition : elle ne se réduit plus au schéma logique Sujet / Attribut (ou Prédicat), reliés par la copule *être*, mais elle est réanalysée en fonctions syntaxiques : le schéma type devient Sujet / Verbe / Objet, le sujet et l'attribut n'étant plus que des fonctions grammaticales parmi d'autres ; les compléments circonstanciels font leur apparition. Le nouveau modèle passe mieux dans les classes, et le maître n'a plus à faire des acrobaties pour retrouver la structure 'Sujet – être – Attribut' dans *Jean mange une pomme*. A côté de la phrase simple (équivalente à la proposition) apparaît la phrase complexe, constituée de plusieurs 'propositions', entendues comme groupes de mots organisés autour d'un verbe à mode personnel : la proposition est devenue une réalité purement grammaticale, qui n'a plus à être la contrepartie d'un jugement :

« Les propositions complétives [= subordonnées] remplissent dans la phrase les mêmes fonctions que remplissent les mots compléments dans la proposition. » (Larousse, *Grammaire complète, syntaxique et littéraire*, 1868, in Chervel, 1981, p. 214)

Cette nouvelle grammaire, dite 'traditionnelle' (mais on pourrait tout aussi bien l'appeler 'grammaire moderne') est encore largement la nôtre.

Texte. Tout cet effort collectif est orienté vers un seul but : savoir rédiger par écrit un texte conforme en tous points aux prescriptions de la grammaire, c'est-à-dire dans lequel le modèle soit réalisé sans distorsion : une série de belles phrases, de préférence avec plusieurs propositions, commençant par une majuscule et se terminant par un point. La phrase devient donc ce qu'on cherche dans un texte écrit ; le texte par excellence, seul digne de ce nom, est celui qui est constitué exclusivement de réalisations irréprochables du modèle. Il se produit donc une identification entre la phrase comme modèle de langue (en puissance) et la phrase réalisée en conformité avec le modèle. « Faites des phrases » est la consigne de base de la pratique scolaire (rédactions corrigées, étude des 'bons auteurs', analyse), où 'phrase' s'entend évidemment comme 'phrase grammaticalement correcte' : tout ce qui ne s'y conforme pas est rejeté, réputé invalide, frappé d'indignité voire même d'inexistence. La notion de 'texte oral' frise l'impensable.

Parmi les normes de production des phrases (à l'écrit) figurent en bonne place les consignes de ponctuation : Chervel en a relevé une mention chez un précurseur en la matière, Lhomond (*Eléments de la grammaire française*, 1780, in Chervel 1981, p. 130) :

« Le point se met à la fin des phrases, quand le sens est entièrement fini. »

Cette mention est reprise dans toutes les grammaires scolaires du 19^{ème} et par exemple dans la grammaire la plus répandue et la plus influente, celle de Noël et Chapsal (*Nouvelle grammaire française*, 1823, in Chervel 1981, p. 130) :

« Le point termine toutes les phrases indépendantes. »

Voici donc la phrase devenue unité de segmentation textuelle d'un texte écrit. Ce nouvel usage a connu la fortune que l'on sait, jusqu'à nos jours : il nous est maintenant encore

naturel de nous repérer dans un texte écrit par rapport aux points (ou autres ponctuations fortes) et de dire que tel paragraphe comporte x phrases, en nous basant sur la ponctuation. Ce mode de repérage dans un texte écrit est usuel, commode, incontournable même dans son ordre : aucun autre procédé n'est disponible pour le concurrencer ou le supplanter.

Et qu'est devenu le **Discours** ? il a plus ou moins disparu, cessé d'être un objet d'étude en lui-même et pour lui-même, tant il s'efface derrière son produit, le Texte. La rhétorique et les figures tendent à disparaître des grammaires. L'histoire montre ainsi combien il est difficile de tenir distinctes les trois phases du langage (le langage en puissance, l'activité de langage et son résultat) : les classiques articulaient la langue et le discours (duquel le texte n'était pas distingué), les grammairiens du 19^{ème} mettaient en rapport la langue et le texte (en court-circuitant le discours).

La question des **unités** est donc réglée : la seule unité est la phrase, pour la langue comme pour le texte. La phrase à elle seule remplace à la fois, du côté de la langue, la proposition, et du côté du discours ou du texte, tout ce que la rhétorique avait pu proposer et particulièrement la période : l'idée - aussi naturelle et tenace qu'impossible à mettre en œuvre avec rigueur - d'un 'sens complet' cimente les deux. On est en somme passé de la consigne d'écriture que donnait Lhomond :

Une phrase (par définition bien formée) se met entre deux points
à une pseudo-constatation (par fausse symétrie) :

Ce qu'il y a entre deux points est une phrase.

Les logiciens apprécieront la validité du raisonnement. : ce qu'il y a entre deux points est présumé être une phrase conforme au modèle grammatical de la langue, alors que chacun sait que ce n'est pas toujours le cas. Mais on se désintéresse des cas de non-conformité au modèle, ou même on les rejette (dénier de réalité). L'oral, bien entendu n'intéresse personne.

On voit que cette notion de phrase, comme appartenant à la fois à la langue (en tant que modèle canonique) et au texte (en tant qu'unité de segmentation) est un idéal contradictoire en lui-même : sa validité supposerait une conformité constante des réalisations au modèle, ce qui n'est pas le cas (réalité reconnue depuis les débuts de la réflexion grammaticale). Les classiques prenaient acte de cette distorsion (avec la rhétorique et les figures, et les unités propres au discours, comme la période), les grammairiens et pédagogues du 19^{ème} l'ont occultée pour des raisons pédagogiques, - et nous avons du mal à revenir de cette illusion, confortable et qui nous fournit en pratique une entrée directe dans les textes écrits (la seule commodément disponible). Et cette conception, idéalisée et fallacieuse, de la phrase fournit de surcroît de nos jours un bon prétexte à ceux qui veulent faire le procès de la phrase, comme s'il n'en existait pas d'autre définition.

2.2.2. Les unités entre langue et discours chez les linguistes (20^{ème} s.)

Face à la promotion du texte écrit et à la mutation théorique concomitante qui vient d'être évoquée, les théoriciens n'ont guère fait entendre leur voix au 19^{ème} siècle : ils étaient occupés prioritairement sur un autre terrain, celui de la reconstruction de l'indo-européen. Ce n'est qu'au début du 20^{ème} siècle, avec Saussure (pourtant indo-européaniste s'il en est !) que les débats théoriques ressurgissent au sujet des distinctions entre la langue et le discours et de leurs unités. L'histoire détaillée des controverses qui n'ont cessé au long du siècle reste à faire. On se contentera ici d'accrocher quelques points essentiels pour notre propos.

Le *Cours de linguistique générale* de Saussure, tel qu'il a été publié après sa mort en 1916, marque les débuts de la linguistique moderne. Certains veulent aujourd'hui opposer le 'vrai' Saussure (dont la pensée aurait été altérée par ses éditeurs posthumes) et le Saussure du

Cours de linguistique. La question n'a toutefois qu'un intérêt historiographique : quelle qu'ait pu être la véritable pensée de Saussure, les idées qui ont joué un rôle déterminant dans l'écllosion de la linguistique comme nouvelle discipline, et influencé en profondeur la recherche pendant plus d'un demi-siècle, sont celles qui sont contenues dans le *Cours*, et ce sont donc ces mêmes idées qui nous intéressent.

Saussure (tel que le *Cours* de 1916 présente sa doctrine) redécouvre l'opposition langue / discours (ou, dans ses termes, langue / parole) :

« Pris dans son tout, le langage est multiforme et hétéroclite. » (éd. 1962, p. 25)

« l'exécution [...] est toujours individuelle, et l'individu en est toujours le maître ; nous l'appellerons la *parole*. » (p. 30)

[La langue est] « un système grammatical existant virtuellement dans chaque cerveau, ou plus exactement dans les cerveaux d'un ensemble d'individus. » (p. 30)

« Sans doute ces deux objets sont étroitement liés et se supposent l'un l'autre [...] Mais tout cela ne les empêche pas d'être deux choses absolument distinctes. » (p.37)

Avec le recul, il apparaît sans doute que ces propos du *Cours* (dont on a fait un des principaux apports théoriques de Saussure) ne font que retrouver des vérités connues de tous, et de longue date, à savoir : les hommes ont dans la tête une langue (langage en puissance), qu'ils utilisent pour produire du discours (ou parole, langage en acte). Mais sans doute aussi faut-il créditer Saussure d'avoir tenu à reformuler, dans un contexte où elles étaient perdues de vue (le contexte de la reconstruction de l'indo-européen et de la linguistique historique) ces vérités d'évidence. Dans ce cadre, le **Texte** n'est même pas mentionné : l'effort de Saussure, grand philologue penché sur des textes, et son mérite, est de chercher à en extraire le système de la langue :

« la linguistique a pour unique et véritable objet la langue envisagée en elle-même et pour elle-même. » (p. 317)

Qu'en est-il alors des **unités** (au-delà du mot), et au premier chef de la phrase ? Force est de constater que Saussure (ou du moins le *Cours*) ne s'intéresse guère au problème, et consacre peu de place à la syntaxe : il fait de la phrase le type de combinaison syntagmatique par excellence, sans en donner de définition précise, et la renvoie au discours (à la parole), sans chercher le moins du monde à en préciser le mode d'existence dans les textes :

« La phrase est le type par excellence du syntagme. Mais elle appartient à la parole, non à la langue. » (p. 172)

On n'en saura pas plus. Curieux paradoxe de Saussure, qui remet sur le devant de la scène l'opposition langue / discours, mais évacue le problème des unités syntaxiques ...

A la suite de Saussure, le courant structuraliste se réclamant de lui réaffirme l'opposition langue / discours (ou langue / parole), mais présente une gamme d'opinions diverses sur le statut des unités et spécialement de la phrase.

Un point en tout cas fait l'unanimité chez les linguistes : tous s'accordent pour rejeter, explicitement ou implicitement, l'idée de la 'grammaire traditionnelle' évoquée au paragraphe précédent (idée passée dans une sorte de 'sens commun', et souvent relayée sans état d'âme par les littéraires), selon laquelle la phrase serait à la fois une unité de langue et une unité de texte écrit (entre deux points). En fait, les linguistes ne se posent guère la question des 'unités du texte'.

Les discussions sont en revanche vives sur le point de savoir si la phrase appartient à la langue ou au discours. Les uns rangent la phrase du côté du discours, d'autres la mettent du

côté de la langue, d'autres encore, soucieux peut-être de ne pas se compromettre, ou non convaincus de la nécessité de ce choix, laissent la question en suspens sans en traiter ...

On se contentera ici de mentionner la position de Benveniste, qui a fait couler beaucoup d'encre :

« Dans le système des signes de la langue, [...] la limite supérieure est tracée par la phrase [...]. Il n'y a pas de phrase hors de la prédication. » (1966, p. 125)

« La phrase est l'unité du discours. » 'Les niveaux de l'analyse linguistique' (1966, p. 130)

A première vue, Benveniste s'inscrit dans une lignée issue de Saussure, et renvoie comme lui la phrase au discours. Mais à y regarder de plus près, les deux points de vue ne sont pas loin de s'opposer : là où Saussure se gardait de toute précision ou définition concernant la phrase, Benveniste fait reposer la phrase sur la prédication (comme il en va pour la proposition chez Aristote ou chez les classiques) et lui assigne la place de signe maximum de la langue. Autant de caractérisations dont il pourrait découler logiquement que la phrase est l'unité maximum de la langue ; dans ces conditions, l'affectation de la phrase au discours devient même surprenante. Ce que veut dire Benveniste (attaché comme on sait à décrire le système formel de l'énonciation, la conversion de la langue en discours) est sans doute que la phrase est l'unité – seuil, le point de passage de la langue au discours. C'est la phrase qui relie la langue et le discours : à ce titre, il aurait sans doute pu tout aussi bien dire qu'elle appartient au deux (ou à un entre-deux insaisissable).

Après la 'révolution saussurienne', la 'révolution chomskyenne' se caractérise, sur bien des points et en particulier pour ce qui concerne la question des unités linguistiques, par... un retour à des thèses proches de celles des classiques : on sait que Chomsky s'inscrit en rupture par rapport à la linguistique de la première moitié du 20^e siècle (structuralisme issu de Saussure, distributionnalisme) et se revendique d'une tradition aux contours larges, qui embrasse la grammaire classique ('cartésienne', c'est-à-dire notamment représentée par Port-Royal) et une bonne partie de la tradition constituée au 19^{ème} :

« Toute l'information dont nous disposons aujourd'hui me semble prouver que le point de vue traditionnel [= antérieur au structuralisme] était en gros fondamentalement correct sur les questions qu'il abordait. » (1965 ; 1971 p. 15)

On connaît les déclarations programmatiques de Chomsky en 1957, dans lesquelles il avance les premiers éléments d'une grammaire générative (*Syntactic Structures*, 1957, p. 111) :

Axiome : Sentence (S)
Règle 1 : S -> NP + VP

La **phrase** (Sentence) est l' 'axiome' de base de la grammaire, avec un statut de point de départ a priori, nécessaire, non falsifiable, - statut proche de celui qu'elle avait dans la grammaire générale). La règle n°1 (la phrase est constituée d'un Groupe nominal et d'un Groupe Verbal) est un décalque de l'articulation en Sujet / Prédicat, reformulée dans des termes purement morpho-syntaxiques, et plus propres à entrer dans une formalisation. Les évolutions ultérieures de la grammaire générative n'ont jamais remis en question le statut de la phrase comme axiome de départ. Chomsky se retrouve donc en accord avec les classiques pour poser à la base un lien fondamental entre le langage et la pensée (même si, bien entendu, la pensée telle que se la représentaient les gens de Port-Royal a peu à voir avec la pensée vue par Chomsky et telle qu'en traitent les sciences cognitives contemporaines), et pour mettre à

la base du langage une unité axiomatique, la phrase, reposant (comme la proposition, dont elle n'est qu'une forme modernisée) sur la prédication.

On connaît également la façon dont Chomsky reformule à sa manière l'opposition entre langue (compétence) et discours (performance). La **compétence** est la maîtrise que les locuteurs – auditeurs ont de leur langue, et que Chomsky se représente et entend décrire sous la forme d'une grammaire générative :

« Le terme technique de 'compétence' désigne la capacité qu'a le locuteur-auditeur idéal d'associer sons et sens en accord strict avec les règles de la langue. » (*The formal nature of Language*, 1967, trad. fr., 1969, p. 126)

« Manifestement, tout sujet parlant une langue a maîtrisé et intériorisé une grammaire générative où se formule sa connaissance de la langue. » (*Aspects of the theory of Syntax*, 1965, trad. fr. 1971, p. 19)

La **performance** est l'activité de langage, exercée par les locuteurs – auditeurs. Elle implique bien entendu la compétence (comme le langage en acte suppose le langage en puissance), mais aussi des capacités cognitives générales et des facteurs situationnels :

« La performance implique aussi bien de nombreux autres facteurs. [...] Des considérations extra-linguistiques, portant sur le locuteur et sur la situation, jouent un rôle fondamental lorsqu'il s'agit de déterminer comment le discours est produit, identifié, compris. [...] La performance linguistique, en outre, est régie par des principes de structure cognitive (par exemple, par les limitations de la mémoire) qui ne sont pas, à proprement parler, des aspects du langage. » (*The formal nature of Language*, 1967, trad. fr., 1969, p. 126)

Et Chomsky insiste, sur la nécessité de séparer l'étude de la langue (compétence) et celle du discours (performance), comme Saussure, en évoquant la possibilité de plusieurs types de 'modèles de performance' :

« la performance et la compétence doivent être radicalement distinguées, pour permettre une étude efficace de l'une ou de l'autre. [...] Des modèles de performance peuvent être construits de nombreuses et différentes façons. » (*The formal nature of Language*, 1967, trad. fr., 1969, p. 178)

Pas plus que chez les classiques, le **texte** n'est mentionné comme une instance existant pour elle-même. Quant aux **unités**, elles ne peuvent être que celles de la langue (de la compétence), c'est-à-dire, à la base, la phrase.

2.2.3. Le choc du texte oral (fin 20^{ème})

La 2^{ème} irruption du texte dans le champ de la linguistique est due à des facteurs externes : les progrès technologiques ont permis à partir du milieu du 20^e siècle de garder une trace des textes oraux (enregistrement sonore), et l'oral perdait ainsi sa volatilité.

La transcription de l'oral a créé une onde de choc parmi les linguistes. Jusque-là, dans l'opinion générale, la langue parlée était considérée globalement comme moins soignée et de moindre qualité que l'expression écrite (ce qui pouvait aisément s'expliquer par leurs conditions respectives d'utilisation), et par suite considérée comme un moins bon représentant de la langue, et moins digne d'intérêt. Bien entendu des considérations sur le statut social des locuteurs censés se limiter à l'expression orale jouaient dans le même sens. Mais on admettait implicitement que l'oral des personnes cultivées, censément dépositaires du meilleur français, devait, pour peu qu'elles fissent attention à leur élocution, être exempt d'imperfections telles que répétitions, redites, rupture de construction, etc., et n'exhiber que des phrases parfaitement formées.

Or la réalité apparut toute différente : l'oral même le plus surveillé (lecture d'un texte écrit mise à part), quand on le transcrit *verbatim*, apparaît truffé de toutes ces imperfections dont on le croyait exempt (on peut s'en rendre compte en relisant l'échantillon présenté en début d'article), et qui avaient totalement échappé à la conscience des locuteurs (y compris des linguistes) avant qu'une indiscrete transcription ne les fasse voir. Le choc fut rude, on tombait de haut ! Ce constat douloureux, renouvelé à chaque expérience, finit par imposer aux linguistes de changer leur regard sur l'oral : force fut d'admettre la présence incontournable, même dans l'expression la plus soignée, des 'déviations', ou 'dysfluences' de l'oral, en tant que partie intégrante de l'expression. Par contrecoup, le parler 'populaire' s'en trouva réhabilité (puisqu'il n'y avait plus lieu de le stigmatiser pour sa prétendue différence) et devint un objet plus digne d'intérêt aux yeux des linguistes.

Mais alors comment décrire ces productions orales, si la norme canonique y est soumise à tant de distorsions ? Ces déviations venaient s'ajouter aux difficultés inévitables de segmentation liées à l'absence de ponctuation : l'oral n'offre pas de signes démarcatifs tels que le point à l'écrit, et la prosodie est loin de fournir les indices ou les méthodes fiables de segmentation qu'on en a espérés et attendus (en dépit des progrès techniques de l'analyse du signal acoustique, - ou peut-être à cause d'eux). Le problème des unités, rapportées de façon pressante à un texte pour la première fois dans l'histoire de la grammaire, prenait brutalement de l'acuité. Et les linguistes ne purent que constater que les phrases, telles qu'ils en décrivaient la syntaxe, et qu'ils pouvaient en retrouver la trace chez les bons auteurs, se retrouvaient difficilement et rarement dans les textes oraux. Ils se trouvèrent alors démunis, à court de théorie et de méthode.

La réaction générale fut de tourner les critiques vers le modèle qu'on avait cherché à retrouver dans les textes sans y parvenir : une pluie de critiques s'abattit sur la phrase, à laquelle on reprocha pêle-mêle de ne pas avoir de définition précise, ou d'avoir une définition aberrante, et/ou de ne s'appliquer qu'à l'écrit, - en tous les cas de ne pas permettre de segmenter les textes oraux.

Voici par exemple ce qu'écrivait en 1997 Claire Blanche-Benveniste (dont les travaux pionniers contribuèrent puissamment à développer les études du français parlé) :

« Cette description [de la langue parlée] ne peut pas se faire entièrement dans le cadre prévu pour l'étude du français dûment orthographié et ponctué : il suffit de penser que la délimitation de la phrase n'a pas de strict équivalent dans la langue parlée. On peut penser que la recherche de nouvelles unités et de nouvelles démarches pourra apporter un renouvellement intéressant des études sur la langue française, aussi bien écrite que parlée. » (*Approches de la langue parlée en français*, 1997, Introduction, p. 3 ; citation à comparer avec l'Introduction de la 2^{ème} édition, citée plus bas)

Claire Blanche-Benveniste oppose ici nettement la langue française écrite (le français « dûment orthographié et ponctué ») et la langue française parlée, et semble considérer que ce qu'on appelle communément la 'grammaire' s'identifie à la grammaire du français écrit. La phrase n'est mentionnée que comme unité de 'délimitation' de l'écrit (conception qui n'a jamais été un 'cadre d'étude' en linguistique), ce qui, par définition, ne peut convenir à l'oral. Sur la base de ce constat, un appel est lancé en direction de 'nouvelles unités' pour remplacer la phrase défaillante.

Voici encore ce qu'écrit en 2012 le Groupe de Fribourg dans la *Grammaire de la période* :

« Il y a, explicite dans de nombreuses définitions et sous-jacent aux pratiques de segmentations usuelles, le postulat que la phrase est une unité douée de complétude à la fois

sémantique, syntaxique, prosodique et orthographique [...]. Maintenant, la question est de savoir si la notion de phrase, ainsi composée, fournit un bon instrument pour segmenter les discours et décrire leur articulation. Et la réponse est *non*. » (*Grammaire de la période*, p. 7 - 8)

La phrase est caractérisée comme une unité à la fois de la langue et du texte, écrit ou oral : vision purement fantasmée (« notion confuse » en disent les auteurs p. 13), qui excède même la vision traditionnelle (limitée à l'écrit) dont on a montré au paragraphe précédent l'émergence dans un contexte pédagogique au 19^{ème}, et souligné le caractère contradictoire, - conception traditionnelle, qui, rappelons-le, n'a jamais été avalisée par les linguistes du 20^{ème} (de qui sont les 'nombreuses définitions' évoquées ?). Demander ensuite à une unité aussi confuse d'assurer la segmentation des textes (y compris oraux) ne peut mener qu'à une réponse négative, mais que penser de la question ?

Et le Groupe de Fribourg, pour trouver des unités de remplacement, veut

« proposer un modèle unifié de la combinatoire discursive, du morphème à la période (paragraphe tonal) » (*Grammaire de la période*, 4^{ème} de couverture)

c'est-à-dire un système unique qui passe par-dessus la frontière entre la langue et le discours, les 'unités significatives' (clauses, identifiables et délimitables comme telles) s'emboîtant linéairement ou s'incorporant dans les 'unités communicatives' (périodes). Mais tout le problème est de savoir si un tel système est concevable et possible, de savoir si on peut annuler la différence entre la puissance et l'acte au point de les décrire dans un même 'modèle unifié', le texte étant découpé linéairement dans une série hiérarchisée de structures, aboutissant à une structuration d'ensemble, dans laquelle chaque élément aurait sa place et son statut.

3. Y a-t-il des unités du texte ?

La question de la segmentation des textes est donc inséparable de la problématique des rapports entre la langue, le discours et le texte (c'est-à-dire les rapports entre le langage en puissance et le langage en acte), comme la réflexion l'indique, et comme l'histoire (brièvement retracée dans les paragraphes précédents) le montre. Mais à quoi peut conduire cette constatation ? Comment aborder le nouvel objectif assigné à la linguistique, la 'segmentation des textes' ?

A cet égard, la position fondamentale des classiques, reprise par Chomsky reste sans doute la ligne de départ la plus robuste, à la fois la plus simple et la plus satisfaisante : la langue (ou compétence) donne le modèle des structures à produire, le discours (ou performance) met en œuvre ces structures dans le temps (en tenant compte de nombreux paramètres extralinguistiques), le texte résultant (qu'il soit fixé ou non, volontairement et consciemment par son auteur, ou artificiellement et fortuitement) est la trace de cet effort de production ; dans le cas de l'oral spontané fixé et transcrit, il y a toutes probabilités pour que le texte contienne à la fois des parties normées où se reflètent les règles de la langue, et des parties reflétant les tâtonnements en temps réel (hésitations, répétitions, corrections, ...) du locuteur. Le statut d'un texte (particulièrement, mais pas uniquement, d'un texte oral transcrit) relève donc d'une analyse à plusieurs niveaux, et on ne saurait demander à des unités de langue de se retrouver nécessairement dans le discours et le texte, c'est-à-dire de se retrouver comme unités de segmentation

Nous emprunterons à Claire Blanche-Benveniste elle-même quelques citations allant dans le sens de ce qui vient d'être dit. En effet sa position définitive ne s'identifie pas à la critique de la phrase qui s'exprimait dans l'Introduction citée plus haut (§ 2.2.3.), et l'ouvrage de 1997 doit être relu à la lumière de l'Introduction de sa réédition de 2010 (et du nouvel ouvrage de 2010). Alors que dans l'Introduction de 1997 deux « cadres » différents lui apparaissaient nécessaires pour l'étude du français écrit et du français parlé, cette position est clairement rejetée en 2010 : les deux relèvent d'une seul et même « cadre syntaxique » :

« Il n'est pas question d'opposer à la grammaire de l'écrit une grammaire qui serait spécifiquement celle de l'oral. **Un seul et même cadre syntaxique peut rendre compte des différents usages de l'écrit comme de l'oral.** L'exception est dans la morphologie, qui, en raison du poids énorme de l'orthographe grammaticale, doit être traitée séparément pour la langue écrite et pour la langue parlée. » (*Approches...*, éd. 2010, p. 11)

Sans doute cette révision ne va-t-elle pas jusqu'à assimiler ce cadre syntaxique fondamental et celui de la phrase (adéquatement redéfinie), mais la démarche de Claire Blanche-Benveniste relève ici davantage d'une stratégie d'évitement ou de contournement de la phrase (ou un souci de ne pas afficher un retournement d'opinion) que d'un rejet (la notion même de phrase, que ce soit pour y recourir, ou la critiquer, est absente de son autre ouvrage de 2010).

En tout état de cause, la clé des différences entre un texte écrit et un texte de langue parlée transcrite est dans le simple fait que

« Les textes de langue parlée sont rarement des produits finis. » (*Approches...*, éd. 2010, p. 26),

ils ne sont que la trace d'une activité de discours, dont Claire Blanche-Benveniste cherche à épingle les manifestations présentes en tant que telles dans le texte :

« Le mode de production de la langue parlée fait que les locuteurs, répètent, hésitent, se corrigent et laissent certaines parties inachevées. On peut y voir autant d'accidents qui trahissent la langue et qui ne méritent pas qu'on s'y attarde. C'est le parti le plus traditionnel. On peut au contraire y voir des indices précieux pour comprendre comment se forment les discours et comment le langage se met en œuvre. » (p. 56)

« Les bribes, hésitations, répétitions amorces ou corrections, si caractéristiques de la mise en place des discours, ne sont pas la syntaxe de la langue. Ce sont des propriétés, fort intéressantes à étudier, des modes de production de la langue parlée. Il paraît essentiel de bien distinguer les deux domaines d'analyse. » (p. 99)

Les formulations de Claire Blanche-Benveniste sont très nettes : les « bribes, hésitations, répétitions, amorces ou corrections », si elles ne sont pas des « accidents qui trahissent la langue », ne reflètent pas pour autant la langue elle-même : elles « ne sont pas la syntaxe de la langue » (expression dont on remarquera la force), elles n'appartiennent pas à la langue, mais sont des indices de la formation du discours, des propriétés du mode de production du discours. Dès lors, la distinction est posée entre deux domaines d'analyse qu'il est « essentiel de bien distinguer » : la syntaxe de la langue d'une part, les modes de production du discours d'autre part.

Cela étant, la nécessité de les distinguer n'entraîne pas que la distinction soit facile : bon nombre de bonnes vieilles 'figures de rhétorique' de l'époque classique sont à nouveau répertoriées, mais la question se pose de les mettre du côté de la langue (comme le propose Claire Blanche Benveniste) du côté du discours :

« Certains phénomènes que l'on classait au début du siècle dans la stylistique, sous les termes de mise en relief, expressivité, ruptures de construction, pléonasmes, ont été intégrés

dans le domaine de la grammaire » (p. 78). Exemples : l'anacolithe ou dislocation (p. 78), les 'figures de dubitation' (p. 108), le zeugme ou projection (p. 187).

Et en ce qui concerne la segmentation des **Textes**, le dernier mot de Claire Blanche-Benveniste, dans son dernier ouvrage de 2010 exclut clairement la possibilité d'y distinguer directement des unités d'un seul niveau :

« Comme le travail de mise au point du discours se fait dans le temps même de la production, la langue parlée garde la trace de toutes les démarches faites par les locuteurs ainsi que la trace de tous les commentaires qu'ils font sur ces démarches. Elle garde aussi trace des procédures d'enrichissement des syntagmes, souvent donnés successivement sous plusieurs versions lexicales, toutes importantes pour la structuration de l'ensemble. *Le résultat est un matériau composite, dans lequel il faut établir différents niveaux avant de chercher à délimiter les unités de texte.* »[souligné par nous] (*Le français, Usages de la langue parlée*, 2010, p. 199 – 200)

Cette conclusion est la suite logique de tout le raisonnement : un texte d'oral spontané transcrit contient des éléments hétérogènes, et tous ne s'inscrivent pas dans la même trame : les traces des modes de production du discours ont dans l'échange langagier un mode d'existence spécifique, particulièrement fugace :

« En participant à des conversations, nous sommes généralement attentifs à ce que les autres veulent dire et nous serions embarrassés d'avoir à répéter littéralement la forme exacte de leurs propos. C'est sans doute pourquoi nous sommes généralement si peu gênés par les répétitions, les hésitations et les redites propres au langage parlé improvisé, que nous percevons à peine (alors que ce sont souvent des phénomènes exaspérants quand on les voit par écrit). » (*Approches...* éd. 2010, p. 17)

Conclusions

Essayons de résumer et de rassembler les points principaux de notre enquête.

1. Le point central est qu'il est vain et sans objet de reprocher à la phrase (- ou à tout autre modèle en langue -) de ne pas être un bon instrument de segmentation des textes et de ne pas s'y donner à voir sans altération : rien ne limite le texte à exhiber son modèle intact, le discours fournit mille raisons et occasions de l'adapter ou de le déformer. Rien n'oblige donc à abandonner sur le fond le paradigme classique, clairvoyant à cet égard (comme l'a bien vu Chomsky), qui distingue la langue (ou compétence) et le discours (ou performance), et remet le texte à sa juste place de trace d'un discours.

Une unité à laquelle on demande d'être unité à la fois de la langue et du texte ne peut être que vide de contenu (réduite à un simple syntagme) ou chimérique.

2. Comme modèle de langue (instrument destiné à permettre des discours) il faut sans doute y regarder à deux fois avant de jeter aux orties la phrase : on pourrait montrer, en poursuivant dans le détail l'étude de l'échantillon de texte oral qui a été esquissée au début de cet article, que le locuteur s'efforce de couler son intention de signification dans un moule qui est celui de la phrase, dont la fonction est d'être une sorte de 'format d'échange' de base pour la communication entre un émetteur et un récepteur. Dans la même ligne d'argumentation, voir nos articles de 2008 (« Phrase, séquence, période ») et de 2011 (« Phrase et intégration textuelle »).

Bien entendu la définition de la phrase est un travail toujours perfectible, à approfondir et enrichir (cf. notre article de 2017) : l'antique et vénérable proposition aristotélicienne, réanalysée en fonctions et étendue à la phrase complexe au 19^{ème} s., augmentée des éléments énonciatifs périphériques à la fin du 20^o (en particulier grâce aux travaux de Claire Blanche-Benveniste) est un trésor collectif toujours susceptible d'enrichissement, et qu'il n'est pas aisé de remplacer.

3. Reste à élaborer un véritable modèle du discours, ou modèle de performance : tâche considérable, encore trop peu entreprise. Nombreux sont certes les travaux qui pourraient s'y inscrire de droit : analyse conversationnelle, recherches pragmatiques et psycho-linguistiques, étude de 'stratégies de réparation' ou de 'routines discursives', et beaucoup d'autres ; mais ils restent dispersés, faute de cadre d'ensemble pour les rassembler et les organiser dans un véritable modèle de performance qui s'articulerait explicitement sur un modèle de compétence.

Une des grandes difficultés d'un tel modèle est assurément de tracer la limite entre la langue et le discours : on peut prédire sans risque d'erreur de nombreux tâtonnements à venir, des controverses, beaucoup de conclusions sans cesse révisées ...

4. En ce qui concerne la segmentation des **textes**, si les points précédents sont exacts, une conclusion s'impose : on ne dispose pas de procédure adéquate, théoriquement et méthodologiquement, pour segmenter des textes, en particulier des textes d'oral transcrit. Dans les termes (déjà cités) de Claire Blanche Benveniste : un texte de langue parlée « est un matériau composite, dans lequel il faut établir différents niveaux avant de chercher à délimiter les unités de texte » (2010, p. 200). L'établissement des différents niveaux restant à faire, la segmentation est (pour le moment du moins) hors d'atteinte. Conclusion négative, qui n'aura pas les faveurs de tous ceux qui sont engagés dans des travaux de segmentation et d'analyse de corpus ...

Mais peut-être est-il possible (perspective évoquée par Sylvain Kahane) de miser sur les capacités d'apprentissage automatique des systèmes informatiques, qui seraient capables, sinon de segmenter les textes, du moins de déterminer automatiquement les phénomènes relevant de la performance, et, les ayant écartés, de construire des modèles de compétence adéquats, utilisables pour engendrer les textes. L'avenir dira si la machine peut faire mieux que l'homme ...

La question des textes et de leur segmentation en unités, en tout cas, ne concerne pas exclusivement les textes oraux : il n'y a que dans la rédaction du bon écolier, appliqué à enchaîner des phrases closes sur elles-mêmes, que la segmentation en phrases, voulue et planifiée, va de soi, si tout va bien. Un texte écrit (livré par son auteur comme 'produit fini') ne se laisse normalement pas décomposer de cette façon (qu'on prenne en compte la ponctuation ou non) : le facteur temps joue dans les textes écrits comme à l'oral, de façon moins contraignante pour l'émetteur sans doute, mais sans pouvoir être évacué : une page de roman, un article de presse ne sont pas segmentables sans difficultés, et posent en définitive les mêmes problèmes qu'une transcription d'oral, de façon moins criante et moins aiguë, sans doute, mais tout aussi réelle. Les segments entre deux points ne correspondent à aucune définition linguistique consistante.

En dernière analyse, on se trouve toujours ramené à la distinction entre la puissance et l'acte, entre la langue (ou compétence), le discours (ou performance), et le texte : toujours ces vérités d'évidence si complexes, ces banalités si subtiles et si embarrassantes... Les apprivoiser peu à peu, dans une lente construction continue, à travers le bruit incessant des

proclamations de rupture et la pression grandissante de la technologie, c'est peut-être aussi en cela que réside les progrès en sciences humaines.

Références

- ARNAULD & LANCELOT, 1969 [1660], *Grammaire générale et raisonnée*, Paris, Paulet
- ARNAULD & NICOLE, 1993 [1662], *La Logique ou l'art de penser*, Paris, Vrin (éd. critique)
- BENVENISTE Émile, 1966, *Eléments de linguistique générale*, T.1, Paris, Gallimard
- BENVENISTE Émile, 1974, *Eléments de linguistique générale*, T.2, Paris, Gallimard
- BLANCHE BENVENISTE Claire, 1997, *Approches de la langue parlée en français*, Paris, Ophrys
- BLANCHE BENVENISTE Claire, 2010 (réédition), *Approches de la langue parlée en français*, Paris, Ophrys
- BLANCHE BENVENISTE Claire, 2010, *Le français. Usages de la langue parlée*, Leuven – Paris, Peeters
- CHERVEL André, 1981, *Histoire de la grammaire scolaire*, Payot
- CHOMSKY Noam, 1957, *Syntactic Structures*, Mouton
- CHOMSKY Noam, 1969, *La linguistique cartésienne*, et *La nature formelle du langage*, Paris, Seuil (trad. de *The formal nature of Language*, 1967)
- CHOMSKY Noam, 1971, *Aspects de la théorie syntaxique*, Paris, Seuil (trad. de *Aspects of the theory of Syntax*, 1965)
- DUMARSAIS, 1988, *Des Tropes* [1730] [suivi de l'article 'Figure'], Paris, Flammarion
- Groupe de Fribourg, 2012, *Grammaire de la période*, Bern, Peter Lang
- LAMY A., 1998 [1675], *Rhétorique*, Paris, PUF (éd. critique)
- LE GOFFIC Pierre, 2008, « Phrase, séquence, période », in Van RAEMDONCK Dan (éd.), *Modèles Syntaxiques. La syntaxe à l'aube du XXIème siècle*, Bruxelles – Berne, Peter Lang, p. 329 - 356
- LE GOFFIC Pierre, 2011, « Phrase et intégration textuelle », *Langue Française*, 170, p. 11 – 28
- LE GOFFIC Pierre, 2017, « A la recherche des unités du langage : une vieille histoire toujours d'actualité », *Langages*, 205, p. 133 - 143
- SAUSSURE (de) Ferdinand, 1962 [1916], *Cours de linguistique générale*, Paris, Payot