

*PEUT-ON PREDIRE LA LONGUEUR D'UNE PHRASE FRANCAISE
LORSQU'ON CONNAIT LA NATURE DE SON PREMIER MOT (1) ?*

ABSTRACT

This paper consists of 2 parts.

§ 1 recalls some facts emphasizing the importance of the initial word cluster of a sentence from the various points of view of psychological linguistics (1.1), grammar (1.2), information theory, stylistics (1.3) and quantitative studies (1.4).

The other §, which analyze a corpus of 2,672 sentences from the Belgian writer Jean Ray's work, make obvious 5 subsets (4.2) of initial words and 23 smaller groups (4.3) each beginning sentences of specific average length.

We conclude (5) that in French one can predict the length of the sentences coming after each particular subset or particular group of initial words. But this prediction, which is merely statistical, is also relative (every mean being situated with reference to the other means).

1. AVANT-PROPOS

1.0. Notre but est de montrer qu'on peut déterminer dans une certaine mesure la longueur d'une phrase française en se fondant sur les caractéristiques de son mot initial. Nous commencerons par rappeler quelques faits qui

mettent en évidence les propriétés privilégiées des premiers termes d'un énoncé.

1.1. Le début d'une phrase soutient l'édifice de celle-ci.

Comme le soulignait D.L. BOLINGER :

"Before the speaker begins, the possibilities of what he will communicate are practically infinite, or, if his utterance is bound within a discourse, they are at least enormously large. When the first word appears, the possibilities are vastly reduced, but that first word has, in communicative value for the hearer, its fullest possible semantic range. The second word follows, narrowing the range, the third comes to narrow it still further, and finally the end is reached at which point the sentence presumably focuses on an event - usually aided by a gesture, a physical context in which only one of several possibilities can be elected, or what-not."
(2)

Tout se passe du reste comme si la venue d'un mot était régie par le(s) mot(s) qui précède(nt). Une idée ayant été engrangée dans la mémoire à court terme (3), le locuteur la communique en choisissant ses mots un à un, non en se conformant à un modèle architectural préalablement conçu.

L. TARNÓCZI résumait récemment la chose :

"J.R. Peirce [*Symboles, signaux et bruits*, Paris, 1966] nous dit par exemple : "Lorsque l'on parle ou que l'on écoute quelqu'un qui parle, on a fortement l'impression que les phrases sont engendrées en grande partie du début à la fin. On a également l'impression que la personne qui engendre une phrase n'a pas de plan très déterminé en tête à n'importe quel moment, mais qu'elle élabore le plan au fur et à mesure qu'elle avance." H. Glinz [*Deutsche Syntax*, Stuttgart, 1965] a remarqué, de sa part, qu'il n'a pu observer les mécanismes décrits par Chomsky ni chez lui, ni chez un locuteur ou un auditeur quelconque.

“En effet, ce qu’un émetteur a en tête, c’est son vocabulaire
“individuel, mis en relation avec des concepts plus ou moins
“individuels, sa grammaire fragmentaire, ainsi que son intention
“communicative, mais point d’axiomes, ni même de règles déduc-
“tionnelles formelles. Lors de la construction de ses messages, il
“progresses par des balbutiements et des gémissements répétés,
“ayant recours aussi à des formes de remplissage automatiques
“(eh bien, alors, voyons, etc., la prononciation étant fortement
“allongée), pour trouver le temps nécessaire à la réflexion. Cette
“réflexion ne porte cependant pas sur des opérations transforma-
“tionnelles ou substitutionnelles, mais sur le choix du mot et de
“la forme suivants.” (4)

Il est assez suggestif à ce propos qu’un fait de parole comme *la personne que je ... dont je parle* soit particulièrement fréquent (malgré la faible capacité d’anticipation requise par le choix correct du pronom relatif) ou qu’en français le participe passé conjugué avec avoir s’accorde en genre et en nombre avec son objet direct exclusivement si cet objet le *précède*.

Cela noté, on peut encore évoquer ici telle constatation de M.D.S. BRAINE concernant l’apprentissage du langage chez les tout jeunes enfants :

“There are two word-classes : pivots and X-words. The pivots
“are few in number; they tend to occur in several word combi-
“nations; and each is associated with a particular utterance
“position. [...] The X-class is a large open class containing
“the child’s entire vocabulary except for some of the pivots.
“[...] Single-word utterances are X. Multiword utterances
“are either P₁X or XP₂, the former being much more fre-
“quent [...] (5)

ou telle autre de M.-M. BRAUN-LAMESCH :

“[...] en présentant oralement à des sujets âgés de 5, 6 et 8 ans
“des mots en désordre et en leur demandant de restructurer les

“énoncés, nous constatons qu’ils prenaient fréquemment un des
“termes les plus importants - très souvent le nom - et qu’ils cons-
“truisaient une phrase autour [...] (6)

1.2. Le début d’une phrase peut imposer à l’édifice de celle-ci un type de structure particulier.

C’est ainsi que l’ordre Sujet-Verbe-Complément d’Objet, dans la plupart des langues où il est régulier, subit une altération à l’intérieur des énoncés qui commencent par un vocable interrogatif. Le même phénomène se produit d’autre part à l’intérieur des phrases débutant par la conjonction *aussi* en français, par un mot qui n’appartient pas au syntagme sujet en néerlandais (7), par *there* ou par un complément négatif en anglais (8), etc.

1.3.0. Le début d’une phrase est l’endroit le plus riche en données de celle-ci.

1.3.1. La Théorie de l’Information a établi que davantage de renseignements se trouvent en général véhiculés par un mot situé dans la première partie d’une phrase que par un mot situé en une autre position (9).

1.3.2. Au demeurant, la stylistique a constaté que les écrivains ont tendance à mettre en tête d’énoncé les unités lexicales qui leur semblent les plus importantes.

Ceci a souvent été signalé non seulement pour les langues où l’ordre des mots est “libre”, comme le latin (10), mais aussi pour les langues où l’ordre des mots est “fixe”, comme la nôtre (11). Le français possède d’ailleurs l’idiotisme *c’est...que*. Il peut recourir à la construction passive, engendrer un nombre considérable de “structures de surface” différentes grâce au jeu

des "permutations", etc. Ajoutons que, comme l'écrivait F. FRANÇOIS :

"De toute façon il importe de se rappeler que la syntaxe n'est pas primitivement une forme obligatoire, mais un moyen de hiérarchiser les éléments du discours et que, s'il croit pouvoir s'en passer, le locuteur n'hésitera pas à le faire, en utilisant par exemple les ruptures de construction ou les "appendices" (les "remords) de toutes les façons possibles, l'indication du lien pouvant être purement sémantique ou marquée par un procédé non syntaxique : pause ou intonation." (12)

Enfin, il est révélateur que beaucoup de stylostatisticiens, notamment L.T. Milic dans son étude sur Swift (13) et R. Cluett dans son étude sur Sprat (14), aient été amenés à faire du premier mot des phrases un objet d'observation privilégié.

1.4. Le début d'une phrase et le reste de celle-ci sont compris dans un même système de relations quantitatives. Que l'on songe, entre autres, au remarquable article de V. Mihăilescu-Urechia dont la thèse est la suivante :

"By using one of the formulae given below [...] one gets directly the final term (F) of the sentence, on the basis of a relation established between the point of maximum height of the partially destroyed sentence (P), N (the number of words), the upper limit of the band in the partially destroyed sentence (L) and the lower limit of the same band (l)." (15)

2. MATERIAUX ANALYSES

2.0. Ces jalons posés, nous allons essayer de montrer à partir d'un corpus de 2672 phrases empruntées à l'oeuvre de Jean Ray qu'en ce qui

concerne le nombre de mots pleins par énoncé chaque S. (abréviation dont nous userons pour désigner une série de phrases commençant par un mot de même nature) se situe d'une manière relativement identifiable par rapport aux autres S.

Ceci appelle certains commentaires.

2.1. Jean Ray est un écrivain belge contemporain (Gand; 1887-1964; de son vrai nom Raymond De Kremer). Notre choix s'est porté sur ses écrits pour deux raisons. D'abord, sa langue qui est très peu "littéraire" s'avère assez proche de la langue parlée. Ensuite, près de la moitié de ses phrases (45,620 % dans notre corpus) sont des phrases de conversation (16) dont la critique s'accorde à souligner le caractère fort peu apprêté (17). Ces considérations ne sont pas sans importance. Ce qui a été dit en 1.1. est en effet principalement valable pour la communication verbale spontanée.

2.2. On discerne nettement quatre tranches dans la biographie littéraire de Jean Ray : 1° avant son incarcération (en 1926), 2° avant la guerre, 3° de 1942 à 1947, 4° avant sa mort. Nous avons sélectionné quatre oeuvres très caractéristiques chacune de la manière d'écrire de Jean Ray durant l'une des périodes biographiques en question. Il s'agit : 1°) des *Contes du Whisky* 2°) des fascicules 69 et 141 de la série *Harry Dickson* 3°) du *Livre des Fantômes* 4°) du *Carrousel des Maléfices* (18). Leurs fragments (19) ont respectivement formé les collections de phrases qui se verront désignées plus bas par les abréviations *CW*, *HD*, *LF* et *CM*.

Nous avons constitué les collections en cause en un amalgame où nous avons calculé, d'une part, la longueur moyenne \bar{X} et l'écart-type σ de

toutes les phrases, d'autre part, les longueurs moyennes \bar{x} des diverses S. de phrases. Nous avons ensuite mis en relief la différence $\bar{X}-\bar{x}$ pour chacune des S. en situant ces dernières par rapport à la séquence des dix classes-repères ci-après :

- séries où \bar{x} est inférieure de plus de $4/6$ de σ à \bar{X} (renseignées par la lettre z en colonne *a* dans notre tableau synoptique);
- séries où \bar{x} est inférieure de plus de $3/6$, $2/6$, $1/6$ ou $0/6$ de σ à \bar{X} (renseignées par d, c, b ou a en *a*);
- séries où \bar{x} est supérieure de plus de $0/6$, $1/6$, $2/6$ ou $3/6$ de σ à \bar{X} (renseignées par A, B, C ou D en *a*);
- séries où \bar{x} est supérieure de plus de $4/6$ de σ à \bar{X} (renseignées par Z en *a*).

Ces chiffres et points de repère ne fournissent à proprement parler que des renseignements relatifs à la langue de Jean Ray. Toutefois, par voie d'extrapolation, il ne sera pas impossible d'en extraire quelques informations concernant la langue française pourvu qu'on veuille bien prendre deux précautions.

Premièrement, on renoncera à exploiter les chiffres obtenus pour les S. qui ne sont pas représentées à la fois dans *CW*, *HD*, *LF* et *CM*. Deuxièmement, on n'exploitera pas non plus les chiffres obtenus pour les S. dont les caractéristiques fluctuent excessivement de l'une à l'autre des quatre collections. A cette fin, nous avons calculé les valeurs \bar{X} , σ et \bar{x} pour chaque collection et nous avons établi les 10 classes-repères correspondantes (exactement de la même manière que nous avons procédé pour l'amalgame *CW-HD-LF-CM*). Nous considérerons comme variant d'une façon prohibitive toute S. située par les diverses collections dans des classes-repères qu'il est impossible de

comprendre à l'intérieur d'un éventail de 6 classes contiguës (ainsi la S. des phrases commençant par un article partitif; en effet, les diverses classes-repères où elle s'inscrit - c, z, B et z - ne peuvent pas être englobées dans un empan inférieur à celui qui couvre les 7 classes contiguës z-d-c-b-a-A-B). De telles S. seront indiquées par un V en colonne β dans le tableau qui suit (les autres se verront affectées à cet endroit d'un chiffre 3, 2 ou 1 selon qu'elles seront situées par les diverses collections dans des classes-repères susceptibles d'être comprises à l'intérieur d'un éventail de 6, 4 ou 2 classes contiguës).

Il est vraisemblable que les caractéristiques des S. pour lesquelles la double exigence qui précède sera satisfaite équivaldront à des traits propres au français plus qu'à des traits propres à Jean Ray. D'une part, en effet, *Les Contes du Whisky*, *Le Livre des Fantômes* et *Le Carrousel des Maléfices* sont des recueils de contes fantastiques assez brefs tandis que les *Harry Dickson* sont constitués de récits policiers relativement longs. Or des collections de textes de longueurs et de genres différents (20) sont censées avoir des caractéristiques quantitatives fort dissemblables. D'autre part, *Les Contes du Whisky* constituent le premier recueil publié par notre auteur, *Le Carrousel des Maléfices* constitue le dernier et l'intervalle entre les deux est de plus de 40 ans. Or, on sait combien la technique d'un écrivain change au cours des années (21). Les \bar{X} et les σ des collections (dont nous faisons état au bas de notre tableau) sont d'ailleurs assez éloquents à cet égard.

2.3. La longueur des phrases est exprimée dans cet article par le nombre de leurs mots "pleins" : les résultats obtenus sont strictement parallèles aux résultats qu'on obtiendrait en recensant tous les agrégats typographiques (22) et la méthode est beaucoup plus économique.

Un mot "plein" a été compté chaque fois que nous avons rencontré un adjectif, un adverbe (même si c'est une onomatopée) ou un nom. Etant donné les réactions psychologiques particulières que suscitent les suites figées de mots (23), nous n'avons également compté que pour un seul mot "plein" certaines expressions stéréotypées. Il s'agit :

- 1°) du mot composé
- 2°) de la série nominale formant un nom de personne, une enseigne d'immeuble, un titre d'oeuvre ou encore une paire nom de rue ou de place-terme équivalant à "rue" ou à "place" en une langue autre que le français
- 3°) de l'intégralité de la forme verbale correspondant à un lemme unique
- 4°) de la locution adverbiale.

Une ségrégation, on le voit, a été établie entre ces deux derniers types de séquences et les locutions adjectives, nominales et verbales qui, d'une façon systématique, ont été décomposées en leurs éléments. Ce distinguo nous semble fondé dans la mesure où, d'une part, les auxiliaires sont seulement ressentis au sein d'une forme verbale comme des monèmes grammaticaux et où, d'autre part, la presque totalité des occurrences de locutions adverbiales (à l'instar, notons-le au passage, des locutions conjonctives et prépositives) provient d'un corpus de lemmes peu nombreux quoique fréquemment sollicité (24). De plus, bon nombre de locutions adverbiales ne comptent pas dans leurs composants un mot de fonction grammaticale analogue à celle du composé (ce qui est loin d'être le cas pour les locutions nominales et verbales). Or, comme le notait F. FRANÇOIS, l'unité d'une locution est plus forte "lorsque les composants ont une autre fonction syntaxique que le composé" (25). Il est du reste suggestif qu'au cours de l'histoire de la langue française une bien plus grande quantité de locutions adverbiales que de locutions adjectives ou verbales se sont vues typographiquement

amalgamées (telles *davantage*, *partout*, *d'arrache-pied*, *sur-le-champ*, etc.).

Signalons enfin que le repérage des lemmes et leur inscription dans la catégorie grammaticale idoine se sont strictement conformés au *Petit Robert* ou, quand ni celui-ci, ni le *Grand Robert* n'étaient suffisamment explicites, au *Grevisse* (8e éd.).

2.4. Par "mot qui commence une phrase", nous entendons *mot de n'importe quel type* (et pas seulement mot "plein").

Les distinctions relatives à la nature de ce mot initial sur lesquelles repose la délimitation de nos séries de phrases sont suffisamment détaillées dans notre tableau pour qu'il soit inutile de les énumérer ici. Sauf peut-être en traitant certaines locutions comme des unités (cf. 2.3.), nous ne pensons pas que, trop poussées, elles aient fait la part trop belle aux capacités discriminantes d'un auditeur ignorant tout de l'énoncé à venir. Les unes qui sont d'ordre sémantique vont de soi, les autres (par exemple l'opposition de "quel homme !" à "quel homme ?") sont autorisées par des considérations phonologiques (26).

En fait, il eût même été souhaitable de constituer davantage de S., mais la population étudiée était trop peu nombreuse pour qu'on procédât de la sorte.

2.5. La phrase est le fait de parole qui, selon S. POTTER

"[...] may be described as that part of chain of discourse
"which occurs between two pauses and which inevitably
"shows some kind of arrangement, construction, or *syntax*."
(27)

Etant donné que la ponctuation rend compte des pauses d'une façon assez satisfaisante (28), nous avons considéré lors de l'établissement de notre corpus que les bornes d'une phrase étaient tout simplement les endroits auxquels ., !, ? ou . . . se trouvaient immédiatement suivis d'une majuscule.

Du reste, cette convention offre de nombreux avantages pratiques (29).

2.6. En ce qui concerne notre tableau, précisons encore deux choses.

D'une part, un * venant immédiatement après la désignation d'une sous-catégorie grammaticale notifie qu'aucune phrase ne commence par un mot relevant de cette dernière.

D'autre part, les indices qui apparaissent en colonne γ signalent dans quels ensembles et sous-ensembles il nous a paru expédient de situer les groupes de mots initiaux déterminant une même S. Ces indices spécifient que les mots en cause participent :

- soit d'une phrase affective (I) au niveau de la proposition au moins (exclamative ou impérative : Ia; interrogative : Ib) ou au niveau du mot au moins (Ic)
- soit d'une phrase énonciative (II) pour laquelle une structure plus complexe que {Sujet, Verbe} est possible (IIa) ou pour laquelle une structure plus complexe que {S, V} est presque certaine (IIb).

2.7. Afin d'en terminer avec ce chapitre, notons que, dans une étude tout à fait parallèle à celle qui vient d'être décrite, nous avons calculé la longueur moyenne de S. dont les trois premiers mots "pleins" offrent la même séquence de catégories grammaticales (les noms propres et les verbes

de subordonnée étant au reste considérés comme respectivement distincts des noms communs et des verbes de principale).

Ce second répertoire d'observations ne sera pas systématiquement analysé. Le critère choisi ici pour constituer des S. est en effet beaucoup trop grossier, car des mots susceptibles de différer les uns des autres tant au point de vue sémantique qu'au point de vue phonologique ou syntaxique se voient confondus sous l'étiquette d'une même catégorie grammaticale (30).

Nous n'en exploiterons que deux données particulièrement significatives (dans les notes 38 et 42) et ce, d'une façon toute subsidiaire.

3. TABLEAU DES RESULTATS

Voir ce tableau en Annexe.

4. INTERPRETATION

4.0. Lorsqu'on examine ce répertoire, quelques observations s'imposent à l'esprit.

4.1. Les S. qui ne satisfont pas à la double exigence formulée en 2.2. s'avèrent relativement peu nombreuses. Il y en a 11 en tout (regroupées dans la partie inférieure du tableau) dont 3 *seulement* (celles de l'article partitif, de l'adjectif qualificatif et du nom commun exprimant un titre sans se trouver en apposition) se voient éliminées à cause d'une variation

de collection à collection prohibitive (31). Les traits caractérisant chacune des S. que nous avons délimitées offrent donc une stabilité fort appréciable.

4.2.0. Chacun des ensembles et sous-ensembles définis en 2.6. a une idiosyncrasie marquée.

4.2.1. A une exception près, en effet, les longueurs moyennes de toutes les S. de phrases affectives (I) sont *inférieures* à celles des S. de phrases énonciatives (II).

De plus, les longueurs moyennes de toutes les S. de phrases affectives "au niveau de la *proposition* au moins" (Ia et Ib) sont inférieures à celles des S. dont la physionomie spécifique est moins accusée (S. des phrases affectives "au niveau du *mot* au moins" (Ic)), tandis que, à une exception près, les longueurs moyennes de toutes les S. de phrases énonciatives "pour lesquelles une structure plus complexe que {S,V} est presque *certaine*" (Ib) sont supérieures, pour leur part, à celles des S. dont la physionomie spécifique est moins accusée (S. des phrases énonciatives "pour lesquelles une structure plus complexe que S,V est *possible*" (IIa)).

Ces dichotomies tiennent sans doute au fait que les propos, selon qu'ils sont affectifs ou énonciatifs, émanent en général de situations linguistiques fortement ou faiblement "vécues", les unes étant plus éloquents en elles-mêmes ou/et exigeant des comportements verbaux très directs, les autres pas (32).

4.2.2. Les deux exceptions que nous avons signalées en esquisant ce système d'oppositions binaires ne mettent nullement en question sa pertinence.

La première provient simplement d'un chevauchement de la S. de Ic qui a la moyenne la plus haute (S. de l'impératif figé et de l'onomatopée) et de

la S. de IIa qui a la moyenne la plus basse. Or (cf. 4.2.1.) Ic est le sous-ensemble de I qui a le moins le caractère "affectif" et IIa est le sous-ensemble de II qui a le moins le caractère "énonciatif".

La seconde exception est engendrée par la moyenne anormalement faible de la S. de la locution prépositive. Cette irrégularité sera expliquée en 4.3.5.

4.2.3. Quant aux rapports réciproques des sous-ensembles Ia et Ib, il n'est pas possible de les résumer en une opposition globale. Contentons-nous de rapprocher les 2 paires de S. qui sont comparables chez eux.

On constate que les phrases commençant par un adjectif exclamatif sont plus courtes que celles commençant par un adjectif interrogatif alors que les phrases commençant par un verbe impératif sont plus longues que celles commençant par un verbe interrogatif.

Cela peut s'expliquer de la façon suivante. Plus le destinataire veut déterminer un comportement consistant chez le destinataire, plus son discours doit se faire circonstancié. Or un adjectif exclamatif ne sollicite guère de réponse précise, un adjectif, un verbe interrogatif sollicite une réponse verbale et un verbe impératif sollicite souvent une réponse motrice.

4.3.0. A l'intérieur de chacun de nos cinq sous-ensembles maintenant, une hiérarchie s'esquisse dont il est assez facile de saisir la logique.

4.3.1. Dans l'un comme dans l'autre des sous-ensembles adjacents Ia et Ib, la longueur moyenne des S. croît, avons-nous l'impression, à mesure que décroissent le pouvoir évocateur et le potentiel d'autonomie syntaxique des mots qui les débudent.

Primo, en effet, les substantifs constituent des mots particulièrement évocateurs (33) et, ceci découlant sans doute de cela, fort peu dépendants (34). Or ce sont les noms communs n'exprimant pas un titre et ne se trouvant pas en apostrophe qui déterminent la moyenne de S. la plus basse dans le sous-ensemble Ia tandis que, parallèlement, les pronoms interrogatifs (lesquels désignent comme les noms des entités précises et sont relativement autonomes) déterminent la moyenne de S. la plus basse dans le sous-ensemble Ib.

Secundo, les adjectifs ont un pouvoir évocateur moindre que les substantifs (35) et supposent régulièrement la présence d'un nom au niveau du syntagme nominal tandis que les verbes ont un pouvoir évocateur moindre que les adjectifs (36) et supposent souvent la présence d'un nom à un niveau supérieur à celui du syntagme nominal. Or, en Ia comme en Ib, après les S. de phrases dont le premier mot désigne une entité, viennent, au point de vue de la longueur moyenne, celles dont le premier mot concerne une qualité (adjectifs exclamatifs et interrogatifs), puis celles dont le premier mot concerne un procès (verbes interrogatifs et impératifs).

Tertio, les adverbes ont un pouvoir évocateur encore moindre que les adjectifs et adverbes et supposent régulièrement la présence d'un mot exprimant une qualité ou un procès. Or la S. qui leur correspond détient la longueur moyenne la plus élevée de tout le sous-ensemble Ib.

4.3.2. Au sein du sous-ensemble Ic, il apparaît que les phrases sont plus ou moins courtes selon que leur mot initial suppose de la part du destinataire une plus ou moins grande assurance à l'endroit de la chose à signifier ou/et vis-à-vis du destinataire.

De fait, les phrases débutant par un adverbe d'opinion (qui implique une représentation nette du propos à tenir) et celles débutant par une apostrophe (qui implique un lien direct avec l'interlocuteur) sont plus brèves que celles débutant par un impératif figé ou une onomatopée, catégories dont les lemmes les plus fréquemment employés (37) impliquent un préalable au propos essentiel (expulsion d'une émotion ou ménagement de l'auditeur).

4.3.3. A l'intérieur du sous-ensemble IIa, la longueur de la phrase et la longueur que son syntagme initial est présumé atteindre - eu égard à son premier mot - pour désigner avec précision un objet, sont corrélatives.

En effet, si l'on classe toutes les S. de l'ensemble en cause (sauf la S. correspondant à la locution adverbiale qui verra son cas examiné en 4.3.5.), on constate que viennent, par ordre progressif de longueur moyenne, les phrases qui commencent par :

- *primo*, un nom propre. Or ce mot, désignant une entité unique, n'a guère besoin d'être déterminé (38).

- *secundo*, une conjonction de coordination. Or ce mot implique, à l'initiale, un lien étroit entre une phrase énoncée et une phrase à énoncer. Il suppose en conséquence, le plus souvent, que l'objet qui va être abordé a déjà été tant soit peu évoqué au travers de l'un ou l'autre de ses aspects, voire même qu'il a déjà été expressément désigné.

- *tertio*, un adjectif possessif. Or la réalité que ce mot suggère est située très économiquement dans la mesure où il la rapporte à une réalité définie dans un contexte antérieur.

- *quarto*, un pronom non interrogatif-non relatif. Or ce mot, s'il implique presque nécessairement la venue d'un verbe, est spécifiquement anaphorique.

- *quinto*, un adjectif non exclamatif-non interrogatif-etc.(39) et
- *sexto*, un article. Or ces mots réclament tous deux la venue d'un substantif (40) (quant à la préséance de la S. de l'un sur les S. de l'autre, elle s'explique par le fait que l'adjectif non exclamatif-non interrogatif-etc., mettant l'accent sur un aspect très accessoire de l'objet du syntagme initial, présage un énoncé sensiblement moins analytique d'esprit que ne le fait l'article).

- *septimo*, un adverbe. Or ce mot ne se rapporte pas à un vocable qui renverrait directement à une réalité précise (cf. 4.3.1., *tertio*).

Cela dit, il est révélateur que trois des 4 S. que leur longueur moyenne place tout au bas de l'échelle dans le sous-ensemble Ia offrent des caractères dont nous avons déjà eu l'occasion de souligner l'incidence sur la brièveté des phrases : une forte connotation affective (cf. 4.2.1.) jointe à fort pouvoir évocateur (cf. 4.3.1.). Il s'agit des S. qui correspondent au nom propre, au pronom et à l'adjectif possessif. Rappelons d'ailleurs que le dénombrement des occurrences de ces trois types de mot sert à établir le score *d'intérêt humain* dans le test de lisibilité que R. FLESCHE a élaboré pour l'anglais et que G. DE LANDSHEERE a adapté au français (41).

4.3.4. En ce qui concerne le sous-ensemble IIb enfin, la préposition et la conjonction de subordination initiales font attendre une organisation de l'énoncé respectivement complexe au niveau propositionnel au moins et complexe au niveau suprapositionnel. Rien d'étonnant donc si la S. débutant par la première a une longueur moyenne inférieure à celle de la S. débutant par la seconde (42) (pour les autres S. de IIb, voir le paragraphe suivant).

4.3.5. Si les S. correspondant aux diverses locutions ont été laissées de

côté en 4.3.3. et en 4.3.4., c'est qu'il était plus expédient de les traiter en une seule analyse.

On constate que les phrases commençant par une locution adverbiale et par une locution prépositive sont respectivement plus brèves que les phrases commençant par un adverbe et par une préposition.

Cela se comprend aisément : les unités lexicales longues fournissent beaucoup plus d'informations que les unités lexicales courtes (43). Par conséquent un énoncé relativement bref après une locution suffit à véhiculer le même nombre de renseignements qu'un énoncé relativement long après un mot no locutionnel.

Toutefois, les phrases commençant par une locution conjonctive de subordination sont bien plus longues que les phrases commençant par une conjonction de subordination. Ici, en effet, la dynamique des contraintes est toute différente. Les locutions conjonctives de subordination, à la différence des autres, interviennent au niveau suprapropositionnel. De plus, en ce qui concerne leurs lemmes les plus fréquemment employés (37), elles se décomposent principalement en mots-outils et en adverbes (cf. 4.3.1., *tertio*). Elles interviennent donc dans des discours procédant d'une démarche fort analytique et, par tant, peu compendieuse.

4.4. Cela noté, nous voudrions encore faire état de deux observations qui n'apparaissent pas dans notre tableau.

D'une part, la longueur moyenne des phrases débutant par un adjectif démonstratif ($n = 23$; $\bar{x} = 13,304$) est sensiblement supérieure à celle des phrases débutant par un pronom démonstratif ($n = 175$; $\bar{x} = 10,388$). Ces chiffres, qui d'ailleurs confirment ce que nous avons avancé en 4.3.1., *secundo* et 4.3.3., *quinto*, sont peut-être à examiner à la lumière d'une hypothèse de V. WATERHOUSE :

"Demonstrative pronouns perhaps mark dependence more
"often than do demonstrative adjectives." (44).

D'autre part, 14 de nos S. se trouvent déterminées par des unités lexicales qui tantôt sont affectées d'un des morphèmes du pluriel, tantôt ne le sont pas. Or les phrases qui débutent par un mot au pluriel ont une longueur moyenne ($n_p = 185$; $\bar{x}_p = 11,032$) supérieure à la longueur moyenne des S. en cause ($n_S = 1640$; $\bar{x}_S = 10,474$). De plus, ce qui est vrai pour la totalité des 14 S. est valable, individuellement, pour 10 d'entre elles (45). Tout se passe donc bien comme s'il était plus ardu d'évoquer une entité multiple qu'une entité simple.

5. CONCLUSION

Eu égard à toutes ces considérations, il semble difficile de ne pas conclure qu'en français chaque type de mot fait attendre la fin de la phrase qu'il débute après un nombre de vocables déterminé dans une perspective statistique (il ne s'agit que de moyennes) et relative (les moyennes ne sont situées que les unes par rapport aux autres). Nous dirons, en d'autres termes, qu'à l'initiale chaque type de mot a un potentiel d'ouverture spécifique.

Resterait à savoir si les types de mot que nous avons définis ou d'autres plus nuancés (46) ne témoignent pas aussi d'un potentiel d'ouverture stable en des positions moins privilégiées que le commencement de la phrase.

Mais ceci dépasse - et de loin - notre propos.

Christian DELCOURT

NOTES

- (1) Le présent article dont les données ont été élaborées au moyen de l'ordinateur IBM 360/20 du Laboratoire d'Analyse statistique des Langues anciennes eût été impossible sans le créateur et le directeur de ce centre remarquablement équipé, M. le Professeur L. DELATTE, qui nous a initié aux recherches quantitatives et qui, depuis plusieurs années déjà, encourage et dirige nos travaux avec infiniment de bienveillance. Qu'il nous soit permis de lui exprimer toute notre gratitude. Nous aimerions également remercier ici M. le Professeur Et. EVRARD, Melle S. GOVAERTS et M. J. DENOOZ dont les réalisations effectuées dans le cadre du L.A.S.L.A. nous ont servi de modèles et dont les conseils nous ont éclairé en plus d'une occasion.
- (2) BOLINGER D.L., *Linear Modification*, in *Publications of the Modern Language Association of America*, 67, 1952, pp. 1117-1144.
- (3) Cf. LACHMAN R. et DOOLING D.J., *Connected Discourse and Random Strings : Effects of Number of Inputs on Recognition and Recall*, in *Journal of Experimental Psychology*, 77, 1968, pp. 517-522 et TAYLOR I., *Content and Structure in Sentence Production*, in *Journal of Verbal Learning and Verbal Behaviour*, 8, 1969, pp. 170-175.
- (4) TARNÓCZI L., *Essai de critique au structuralisme*, dans *Linguistics*, 57, 1970, pp. 60-92. Cf. aussi JOHNSON-LAIRD P.N., *The Perception and Memory of Sentences*, in LYONS J. (éd.), *New Horizons in Linguistics*, 3e tir., Penguin Books, 1972, pp. 261-270 et, *mutatis*

mutandis, BRAUN-LAMESCH M.-M., *La compréhension du langage par l'enfant*, Paris, P.U.F., 1972 où on lit ceci (p.96) : "La phrase n'est sans doute pas composée mentalement avant d'être prononcée; beaucoup d'observations montrent d'ailleurs que le comportement verbal en général ne se déroule pas selon un plan clairement conçu à l'avance : les énoncés se construisent au fur et à mesure."

- (5) BRAINE M.D.S., *The Ontogeny of English Phrase Structure : The First Phase*, in *Language*, 39, 1963, pp. 1-13. C'est nous qui soulignons. Le phénomène en cause a également été décrit dans MILLER W. et ERVIN S., *The Development of Grammar in Child Language*, in BELLUGI U. et BROWN R. (éd.), *The Acquisition of Language*, Lafayette, Child Development Publications, 1964 où le terme *d'operator* est employé à la place de celui de *pivot*.
- (6) BRAUN-LAMESCH M.-M., *Op. cit.*, p. 184.
- (7) Il y a du reste de nombreuses exceptions et il n'est pas sans intérêt de souligner la précision des distinctions établies en la matière par la langue, ainsi, l'inversion n'a pas lieu après une proposition subordonnée introduite par *al* ou *hetzij*, ni après une proposition conditionnelle sans préposition, ni après une dizaine de conjonctions de coordination (*en, maar, doch, of(wel), noch, noch ... noch, (wel)nu* et *want*) : cf. par exemple TAVERNIER J.-J., *Nouvelle grammaire-index de la langue néerlandaise*, Bruxelles, A. De Boeck, 1955, pp. 35-35'.
- (8) Ce n'est pas vrai dans tous les cas. Pour un examen détaillé de la question, voir ZANDVOORT R.W., *Grammaire descriptive de*

l'anglais contemporain, trad., Lyon, IAC, 1949, §§ 668-722.

- (9) Cf., entre autres, FORSTER K.I., *Left-to-right Processes in the Construction of Sentences*, in *Journal of Verbal Learning and Verbal Behaviour*, 5, 1966, pp. 285-291, *Sentence Completion Latencies as a Function of Constituent Structure*, in *Journal of Verbal Learning and Verbal Behaviour*, 6, 1967, pp. 878-883 et MARKS L., *Judgements of Grammaticalness of Some English Sentences and Semi-sentences*, in *American Journal of Psychology*, 80, 1967, pp. 196-204.
- (10) Notamment dans MAROUZEAU J., *Traité de Stylistique latine*, 2e éd., Paris, Les Belles Lettres, 1946, p. 324 et WOLFF Ph., *Les origines linguistiques de l'Europe occidentale*, Paris, Hachette, 1970, p. 31. Cf., similairement, GRAPPIN H., *Grammaire de la langue polonaise*, Paris, Droz, 1942, p. 294, etc.
- (11) Cf. BOURDON B., *L'expression des émotions et des tendances dans le langage*, Paris, Alcan, 1892, p. 205 : "Dans ses observations sur la syntaxe de l'ancienne phrase sanscrite et sur la syntaxe grecque, Delbrück [*Die altindische Wortfolge; Die Grundlegen der griechischen Syntax*] arrive également à la conclusion 'qu'un mot se rapproche du commencement d'une phrase ou se place au commencement, lorsqu'il exprime une idée importante'. Même en français, la tendance à placer en tête le mot emphatique réussit, avec plus ou moins de facilité à triompher d'une syntaxe fortement entrée pourtant dans les moeurs, savoir la syntaxe sujet-prédicat".
- (12) FRANÇOIS F., *La description linguistique* et MARTINET A., (éd.), *Le langage*, Paris, Gallimard, 1968, pp. 171-282.

- (13) MILIC L.T., *Unconscious Ordering in the Prose of Swift*, in LEED J., (éd.), *The Computer & Literary Style*, Kent, Kent University Press, 1966, pp. 79-106.
- (14) CLUETT R., *Style, Percept, Personality : A Test Case (Thomas Sprat, 1635-1713)*, in *Computers and the Humanities*, 5, 1971, pp. 257-277.
- (15) MIHĂILESCU-URECHIA V., *Are Novelists Free to Choose their Own Style ?*, in *Linguistics*, 59, 1970, pp. 37-61.
- (16) Précisons de surcroît que n'entrent nullement en compte ici les phrases proférées par le narrateur de récits bâtis selon ces points de vue que recouvrent les étiquettes "I as Witness" ou "I as Protagonist" dans FRIEDMAN N., *Point of View in Fiction : the Development of a Critical Concept*, in *Publications of the Modern Language Association of America*, 1955, pp. 1160-1185.
- (17) Cf., à titre indicatif, BOVEN J., *Spoken op de ruwe heide*, in *Cahiers Jean Ray*, 2, 1972, pp. 16-31 : "Al zijn dialogen munten uit door een frisheid en natuurlijkheid die verder alleen bij de allergrootste auteurs is terug te vinden en die een bewijs leveren voor de taalscheppende kracht van John Flanders [un autre pseudonyme de Raymond De Kremer]".
- (18) Nous avons effectué le codage de ces quatre oeuvres sur la base des dernières éditions susceptibles d'avoir été relues et corrigées par Jean Ray durant la tranche biographique où il en a composé le texte, à savoir : *Les Contes du Whisky*, 2e éd., Paris, La Renaissance du Livre, 15 juin 1925; *L'étrange lueur verte*, Amsterdam,

Roman-, Boek- en Kunsthandel, série "Harry Dickson", n° 69, 1er juillet 1934 et *L'étoile à sept branches*, Amsterdam, Roman-, Boek- en Kunsthandel, série "Harry Dickson", n° 141, 1er juillet 1937; *Le Livre des Fantômes*, Bruxelles-Paris, éd. La Sixaine, 15 mars 1947; *Le Carrousel des Maléfices*, Verviers, Gérard et C°, 1964.

- (19) Nous avons uniquement mis sur fiches les blocs de 15 phrases consécutives qui, dans les éditions décrites à la note précédente, commencent avec les 1ère, 61ème, 121ème, ... phrases de chaque récit indépendant. Deux remarques subsidiaires :
- 1) les blocs codés ont moins de 15 phrases quand, par hasard, le récit s'arrête au milieu de l'un d'eux;
 - 2) le texte perforé pour les *Harry Dickson* est celui de blocs prélevés dans les éditions de 34 et de 37, mais le repérage des 61ème, 121ème, ... phrases, à la suite d'une imprévoyance, a été fait sur la base de la réédition de la série en cause à Verviers, chez Gérard et C°, en 1966 (fasc. 69) et 1968 (fasc. 141).
- (20) Cf. à ce propos KUČERA H. et FRANOIS W.N., *Computational Analysis of Present-Day American English*, Providence, R.I., 1967.
- (21) Cf., entre autres, HERDAN G., *The Advanced Theory of Language as Choice and Chance*, Berlin, Springer-Verlag, 1966, p. 230.
- (22) Chez Jean Ray, il suffit de multiplier par 1,8 (CW) ou 1,7 (chacune des trois autres collections) le nombre de mots "pleins" pour obtenir le nombre d'agrégats typographiques tels qu'ils se voient définis dans DE LANDSHEERE G., *Introduction à la recherche pédagogique*,

2e éd., Liège, G. Thone, 1966, p. 157. Cette affirmation se fonde sur un double codage de 47 phrases de *CW*, 45 de *HD*, 45 de *LF* et 58 de *CM*.

- (23) On voit, par exemple, dans BLUMENTHAL A.L., *Promoted Recall of Sentences*, in *Journal of Verbal Learning and Verbal Behaviour*, 6, 1967, pp. 203-206 qu'une phrase du genre *Gloves were made by hand* est restituée à l'énoncé du syntagme final dans 3,9 dixièmes des cas alors qu'une phrase comme *Gloves were made by tailors* l'est dans 7,2 dixièmes des cas ($p = 0,004$). Cf. aussi BALLY Ch., *Traité de stylistique française*, 3e éd., Paris-Genève, Klincksieck-Georg et Cie, 1951, Vol. I, §§ 79 et 80.
- (24) Cf. BALLY Ch., *Op. cit.*, § 88 : "si en outre l'unité est d'un usage très fréquent, il est impossible de ne pas reconnaître que le groupe équivaut à un mot unique. C'est le cas par exemple de locutions adverbiales comme *tout à l'heure*, *tout de suite*, *à peu près*, *tout à fait*, *sans doute*, *sans cesse*, etc. C'est par ces groupes qu'il faut commencer; ensuite, on étudiera les unités verbales, qui sont extrêmement nombreuses".
- (25) FRANÇOIS F., *Op. cit.*
- (26) Cf. BALLY Ch., *Op. cit.*, Vol. I, § 264, Vol. II, même §, MALMBERG B., *Les domaines de la phonétique*, trad., Paris, P.U.F., 1971, p. 202, etc.
- (27) POTTER S., *Language in the Modern World*, 2e éd., 5ème tir., Penguin Books, 1971, p. 78.

- (28) Cf. par exemple MALMBERG B., *Op. cit.*, p. 205.
- (29) Cf. à ce sujet DELATTE L., *A propos de la mesure de la longueur de phrases dans les textes classiques*, in *Revue de l'Organisation internationale pour l'Etude des Langues anciennes par Ordinateur*, 1966/1, pp. 51-65. Voir aussi DE LANDSHEERE G., *Op. cit.*, p. 158.
- (30) Soit dit au passage, ce répertoire est quand même susceptible de fournir, en marge de notre propos, quelques renseignements qui ne sont pas totalement dépourvus d'intérêt. Il peut notamment donner une idée de la différence qui existe, au point de vue de la séquence des catégories grammaticales, entre les chaînes premier mot - deuxième mot de phrase dans notre corpus et les paires stimulus-réponse dans des épreuves d'association verbale. Quoique nos chaînes se rapprochent des paires dans la mesure où elles sont indépendantes à gauche, les différences sont considérables. En effet, notre répertoire recense, d'une part, 128 et 94 occurrences pour les séquences nom commun - nom commun et nom commun - adjectif, d'autre part, 246 et 35 occurrences pour les séquences adjectif-nom commun et adjectif-adjectif. Par contre, pour les mêmes paires, la compilation de S. WOODWORTH (*Psychologie expérimentale*, trad., 4ème éd., Paris, P.U.F., 1949, ch. XV) enregistre, d'une part, 70 et 20 % des occurrences, d'autre part, 45 et 30 % des occurrences.
- (31) Notons que l'indice V qui apparaît en colonne β pour la première et la dernière de ces 3 S. n'est chaque fois dû qu'à un seul chiffre : celui fourni respectivement par *LF* et *HD*. Dans le cas de la S. de

l'article partitif, ce chiffre discordant tient à une phrase unique. Sa signification est donc très limitée. Dans l'autre cas, il traduit un phénomène curieux : lorsque, pensons-nous, un auteur fait accompagner d'un titre le nom de ses personnages, il a souvent le dessein de créer un climat de solennité ou de componction (ce à quoi contribue ordinairement *d'une façon concomitante* l'allongement des phrases). Or, Jean Ray qui a utilisé le procédé dans cette intention en *CW*, *LF* et *CM* y a recouru à une tout autre fin en *HD* (où 9 titres sur 13 sont des occurrences de *lady*, *lord*, *mister* ou *mistress*) : celle de créer un climat anglais. Quant à la S. de l'adjectif qualificatif, son manque d'unité s'explique vraisemblablement par le fait que la présence en tête de phrase du mot qui la définit peut procéder de constructions fort diverses : d'une part, de tournures comme "Délicieux, murmura Mademoiselle Mathilde [...]", d'autre part, des tournures comme "Horrible fantôme, épargnez-moi [...]" ou encore "Fidèle à ses caprices, le feu du ciel avait épargné la perruche [...]".

- (32) Cf. par exemple BOLINGER D.L., *Op. cit.* : "Commands are typically short. We are not likely to combine two orders in one", etc.
- (33) Cf. tout particulièrement BICKLEY A.C., WEAVER W.W. et FORD F.G., *Information Removed from Multiple Choice Item Responses by Selected Grammatical Categories*, in *Psychological Reports*, 23, 1968, pp. 673-712 : la catégorie grammaticale dont la suppression ampute un texte du plus grand nombre d'informations est celle du substantif.
- (34) Cf. une constatation suggestive dans ARLINSKY M. et EPSTEIN W., *The Frequency of Grammatical Classes on the Production of*

Verbal Strings under Two Responses Sets, in *Psychonomic Science*, 3, 1965, pp. 555-556 : les occurrences de substantifs constituent 68 % des occurrences totales dans les phrases *non* structurées, alors que, dans les phrases structurées, elles constituent 25 % des occurrences seulement.

- (35) Cf. PAIVIO A., *Learning of Adjective-Noun Word Order and Noun Abstractness*, in *Canadian Journal of Psychology*, 17, 1963, pp. 370-379 et KUSYZYN I. et PAIVIO A., *Transition Probability, Word Order, and Noun Abstractness in the Learning of Adjective-Noun Paired Associates*, in *Journal of Experimental Psychology*, 71, 1966, pp. 800-805 : on retient plus facilement une paire nom-adjectif qu'une paire adjectif-nom.
- (36) Cf. GALLAGHER J.W., *Semantic Relationship and the Learning of Syntactic Word Pairs in Children*, in *Journal of Experimental Child Psychology*, 8, 1969, pp. 411-417 et GUPTON T. et FRINCKE G., *Imagery, Mediatonal Instructions, and Noun Position in Free Recall of Noun-Verb Pairs*, in *Journal of Experimental Psychology*, 86, 1970, pp. 461-462 : on retient plus facilement une paire nom-verbe ou une paire adjectif-nom qu'une paire verbe-nom.
- (37) A la fois d'après les chiffres de notre corpus et d'après ceux de VANDER BEKE G.E., *French Word Book*, New York, The Macmillan Company, 1930.
- (38) Notre second répertoire d'observations (cf. 3.6.) fournit quelques données révélatrices à ce sujet : les longueurs moyennes des phrases débutant par les suites nom propre-y-z ($n = 206$; $\bar{x} = 9,252$),

y-nom propre-z (n = 188; \bar{x} = 10,553) et y-z-nom propre (n = 116; \bar{x} = 10,265) sont inférieures aux longueurs moyennes des phrases débutant respectivement par les suites nom commun-y-z (n = 588; \bar{x} = 10,782), y-nom commun-z (n = 741; \bar{x} = 10,819) et y-z-nom commun (n = 749; \bar{x} = 11,596).

- (39) Il faut noter que cette S. ne se voit attribuer en β un indice 3 au lieu d'un indice 2 qu'à 59/5482 de σ près !
- (40) Cf. par exemple JOHNSON-LAIRD P.N., *Op. cit.* : "The initial occurrence of *the* is bound to suggest that a noun will follow later to complete the noun phrase and, ultimately, a verb phrase to complete the sentence. Such predictions are almost certainly set up by the parsing mechanism and tested against the incoming sentence".
- (41) Voir DE LANDSHEERE G., *Op. cit.*, pp. 154-160.
- (42) Ceci est également sensible dans notre second répertoire. Il suffit de comparer, d'une part, les chiffres obtenus pour les phrases commençant par une chaîne verbe de principale-verbe de principale (n = 55; \bar{x} = 8,901) ou nom commun-verbe de principale (n = 179; \bar{x} = 9,910) et, d'autre part, les chiffres obtenus pour les phrases commençant par une chaîne verbe de principale-verbe de subordonnée/verbe de subordonnée-verbe de principale (n = 118; \bar{x} = 9,008/ n = 20; \bar{x} = 9,750) ou nom commun-verbe de principale (n = 179; \bar{x} = 9,910).
- (43) Cf. ZIPF G.K., *The Psycho-Biology of Language*, London, Routledge, 1936.

- (44) WATERHOUSE V., *Independent and Dependent Sentences*, in *International Journal of American Linguistics*, 29, 1963, pp. 45-54.
- (45) Il s'agit des S. de phrases commençant par un adjectif interrogatif (np = 4; $\bar{x}p = 6,285$), un adjectif possessif (np = 15; $\bar{x}p = 12,400$), un pronom non interrogatif-non relatif* (np = 47; $\bar{x}p = 10,595$), un adjectif non exclamatif-non interrogatif-etc. (np = 12; $\bar{x}p = 12,416$), un article défini (np = 50; $\bar{x}p = 12,600$), un article indéfini (np = 16; $\bar{x}p = 12,875$), un verbe de principale non impératif-non interrogatif (np = 1; $\bar{x}p = 16,000$), un adjectif qualificatif (np = 1; $\bar{x}p = 14,000$), un verbe au participe (np = 2; $\bar{x}p = 19,000$) ou un article contracté (np = 3; $\bar{x}p = 18,666$). Les quatre autres S. regroupent les phrases commençant par un nom commun n'exprimant pas un titre et ne se trouvant pas en apposition (np = 4; $\bar{x}p = 3,250$), un verbe interrogatif (np = 2; $\bar{x}p = 5,500$), un verbe impératif (np = 15; $\bar{x}p = 7,333$) ou un impératif figé, une onomatopée (np = 13; $\bar{x}p = 6,461$).
- (46) On pourrait, par exemple, distinguer :
- 1.- les ouvreurs sans potentiel propre : "remplisseurs de pause", incises, etc.
 - 2.- les ouvreurs simples : substantifs, etc.
 - 3.- les ouvreurs décomposables : formes verbales à auxiliaire(s), etc. (cf., en néerlandais, les adverbes pronominaux, les verbes à particule séparable, ...) et
 - 4.- les ouvreurs corrélatifs : d'une part... d'autre part..., non seulement... mais encore..., etc. Des considérations d'ordre prosodique devraient peut-être intervenir également (cf. ce passage

suggestif dans MALMBERG B., *Op. cit.*, p. 203 : “[...] *quand il pleut, l’herbe pousse* [...] La première des propositions est prononcée avec une ‘accentuation’ qui indique qu’elle est incomplète, inachevée, c’est-à-dire, qu’on attend une suite ou qu’une suite est supposée”); etc.

A N N E X E

Nature du mot initial déterminant une S.	nombre de phrases (n) et longueur moyenne de celles-ci (\bar{x} , \bar{X}) à l'intérieur de											
	α	CW, HD, LF, CM	β	γ	α	CW	α	HD	α	LF	α	CM
nom commun ~ titre ~ en apostrophe	d	55 5,781	3	la	z	12 2,666	c	19 6,578	b	10 9,600	z	14 4,642
pron. interrogatif	d	26 5,807	2	lb	d	6 5,833	z	8 4,625	c	7 8,571	z	5 3,800
adj. exclamatif	d	7 5,857	3	la	d	2 6,000	z	1 2,000	a	2 11,500	z	2 2,000
adj. interrogatif	d	7 6,285	3	lb	z	1 3,000	c	3 6,666	b	1 10,000	z	2 5,500
verbe interrogatif	c	31 7,645	3	lb	b	11 9,727	d	5 6,400	z	10 5,800	b	5 8,000
adv. interrogatif	b	27 8,037	3	lb	a	10 10,700	d	4 6,000	z	4 3,750	b	9 7,888
verbe impératif ~ figé	b	66 8,181	2	la	a	15 11,066	c	29 7,275	c	6 9,000	c	16 6,812
adv. d'opinion tonique	b	64 8,437	3	lc	b	19 8,894	A	20 9,600	c	13 7,307	c	12 7,000
nom commun ou propre en apostrophe	b	46 8,456	2	lc	b	21 9,285	c	5 6,800	a	9 10,777	d	11 5,727
nom propre ~ en apostrophe	a	146 9,143	1	lla	a	22 9,772	a	40 8,225	a	38 11,236	b	46 7,913
impér. figé ou onomatopée	a	96 9,187	2	lc	a	36 10,916	b	20 7,600	b	19 10,368	c	21 6,666
locution prépositive	a	19 9,631	1	llb	a	4 10,500	b	2 8,000	b	6 10,333	a	7 9,000
conjonction de coordination	a	199 9,733	2	lla	A	71 11,830	a	41 8,512	b	36 10,000	b	51 7,607
adjectif possessif	A	62 10,322	2	lla	b	19 9,526	a	7 8,857	A	22 12,500	a	14 8,714
pronom ~ interr. ~ relatif*	A	802 10,380	2	lla	A	195 11,620	A	210 9,380	a	165 11,581	a	232 9,387
locution adv. ~ exclam.* ~ interr.	A	88 10,522	2	lla	a	28 10,535	A	19 9,789	a	21 11,407	B	20 10,650
adj. ~ exclam. ~ interr. ~ ordin.* ~ poss. ~ qual. ~ relat.*	A	75 10,800	3	lla	A	25 11,320	C	14 11,071	b	15 10,333	A	21 10,333
article défini	A	291 11,216	2	lla	b	53 9,660	B	83 10,433	B	71 13,281	B	84 11,226
article indéfini	A	97 11,216	3	lla	B	23 13,000	a	32 8,906	A	14 12,428	C	28 11,785
adv. ~ exclam. ~ interr. ~ d'opinion tonique	B	152 11,473	2	lla	B	35 12,657	C	38 12,303	A	33 12,303	A	46 9,652
préposition	B	128 12,179	3	llb	C	18 14,888	a	30 8,666	D	26 15,346	C	54 11,703
conjonction de subordination	B	49 12,489	2	llb	A	14 12,214	C	11 11,636	D	8 16,250	B	16 11,437
locution conjonctive de subordination	B	19 14,263	3	llb	Z	2 21,500	B	8 11,625	Z	2 20,000	Z	7 13,571
article partitif	z	8 5,250	V	lla	c	1 8,000	z	2 1,500	B	1 13,000	z	4 4,500
locution conj. de coordination	d	2 6,000	2	lla	-	-	z	1 5,000	-	-	c	1 7,000
verbe à l'infinitif	b	3 8,333	V	llb	-	-	x	1 5,000	D	1 16,000	z	1 4,000
verbe de princ. ~ impér. ~ interr.	b	2 9,000	V	lla	z	1 2,000	-	-	D	1 16,000	-	-
adjectif qualificatif	a	26 9,961	V	lla	b	9 9,111	C	5 11,200	Z	3 20,000	c	9 6,777
verbe au partic.	a	8 10,250	1	llb	b	1 9,000	b	2 8,000	a	5 11,400	-	-
locution adv. interr.	A	2 11,500	2	lb	a	1 11,000	-	-	-	-	B	1 11,000
nom commun titre ~ en apostrophe	B	37 12,000	V	lla	Z	4 16,750	b	13 8,230	B	6 13,333	Z	14 13,571
verbe au gérondif	B	7 12,000	V	llb	b	2 8,500	Z	1 13,000	-	-	Z	4 13,500
article contracté	D	22 14,136	3	llb	Z	4 24,000	-	-	Z	4 17,000	B	14 10,500
adv. exclamatif	Z	3 15,000	V	la	c	1 8,000	Z	1 17,000	-	-	Z	1 20,000
toute catégorie grammaticale		2672 10,253 $\sigma = 6,780$				666 11,076 $\sigma = 8,017$		675 9,186 $\sigma = 5,482$		559 11,679 $\sigma = 7,124$		772 9,443 $\sigma = 6,078$