

Recherche sur la syntaxe et l'ordre des mots en grec

Bien que le but de mes recherches soit avant tout l'étude de l'ordre des mots dans la phrase grecque classique, il est encore trop tôt pour donner des résultats à ce sujet, étant donné que les travaux de préparation viennent seulement d'être terminés en juillet.(1)

Ceci dit, ma communication se divisera en deux parties : j'esquisserai brièvement la méthode suivie pour aborder ces problèmes en machines; ensuite je montrerai l'intérêt que l'étude computationnelle des phénomènes syntaxiques semble pouvoir présenter d'après les résultats des tout premiers examens.

*

* *

Dans les numéros 1966/4 et 1967/1 de la *Revue de l'Organisation Internationale pour l'Etude des Langues anciennes par Ordinateur*, j'ai proposé un code pour l'analyse morphologique(2) du grec, inspiré de celui du L.A.S.L.A. pour le latin.(3) J'y ai ajouté un code pour l'analyse syntaxique et une méthode pour noter la place des mots dans la phrase.

Qu'il me soit permis d'en rappeler l'essentiel.

La carte-mot porte, à côté de la forme, du lemme et de la référence, les indications suivantes :

- a. colonnes 54-55 : catégorie grammaticale
- 56-59 : analyse morphologique
- 60 : analyse syntaxique
- b. colonnes 72-75 : numéro d'ordre dans l'oeuvre
- 62-65 : numéro d'ordre du "régissant"⁽⁴⁾
- 66 : ponctuation⁽⁵⁾

Les indications sous a. sont notées sous forme de codes, dont voici quelques exemples :

- col. 54 : 1 Substantifs
- 2 Adjectifs
- 3 Numéraux
- 4 Pronoms et adjectifs pronominaux
- 5 Verbes
- 6 Adverbes
- 7 Prépositions
- 8 Négations, Conjonctions, Particules
- 9 Articles

- col. 55 : sert à distinguer des sous-catégories, p. ex.
- sous 4 : les pronoms personnels, démonstratifs, relatifs, etc.
- sous 8 : les conjonctions de coordination, les conjonctions de subordination, les particules de modalité, etc.

- col. 60 : 1 Sujet
2 Accord
3 Objet "direct"
4 Infinitif ou phrase complétive
5 Objet "secondaire"
6 Complément déterminatif ou circonstanciel
7 Mot "connecté"(6)
8 "Complément d'attitude"(7)
9 Coordination

Les indications sous b. permettent d'étudier non seulement la place "absolue" de chaque mot, mais également la place "relative" de différents mots entre eux; plus particulièrement elles permettent d'étudier la place de chaque mot par rapport à son régissant et de tenir compte en même temps des mots qui se trouvent avant, après ou entre les deux.

J'ai déjà dit que je n'en suis pas encore à exploiter toutes les possibilités de ce système. Il a déjà permis la confection en ordinateur d'un nouveau fichier qui donne pour chaque mot des renseignements concernant jusqu'à quatre mots différents en relation syntaxique hiérarchique :

- I. Du mot en question : forme et référence
 catégorie grammaticale
 analyse morphologique et syntaxique
 numéro d'ordre dans la phrase
 ponctuation éventuelle
- II. De son régissant : forme
 catégorie grammaticale
 analyse morphologique et syntaxique
 distance avant ou après le mot I
 ponctuation éventuelle

- III. Du régissant de II : catégorie grammaticale
 analyse morphologique et syntaxique
 distance avant ou après I
 ponctuation éventuelle
- IV. Du régissant de III : catégorie grammaticale

Il est évident qu'un tel fichier peut répondre immédiatement à pas mal de questions qui concernent soit la syntaxe des cas, des modes et des temps, soit la juxtaposition ou la disjonction des mots syntaxiquement liés.

*

* *

Hormis l'utilité pédagogique incontestable d'une grammaire de base, on peut se demander si l'étude computationnelle des phénomènes morphologiques, syntaxiques et stylistiques peut servir à mieux connaître et à décrire plus exactement et la langue en général, et le style des auteurs.

En effet, depuis l'introduction des ordinateurs dans la recherche philologique, on a utilisé abondamment et avec fruit les méthodes statistiques dans les études de phonologie et de lexicologie. La critique littéraire a d'ailleurs employé depuis toujours des catégories de fréquence quand elle parlait de la couleur sonore d'une oeuvre poétique, ou de la richesse du vocabulaire d'un auteur. Les multiples études de la distribution des catégories grammaticales peuvent également se réclamer de la stylistique traditionnelle selon laquelle une préférence pour les verbes est censée caractériser un style dynamique, tandis qu'une abondance d'adjectifs devrait rendre le style pittoresque. Enfin la longueur des phrases, elle aussi objet de plusieurs études statistiques, peut être mise en rapport avec les

catégories de style concis ou prolix. Que peut nous apprendre alors une pareille étude, non plus des sons ou des mots, mais des formes et des fonctions ?

A l'intention de ce colloque j'ai fait deux petites explorations assez superficielles qui, je crois, n'ont quand même pas été sans résultats.

La première est partie d'une constatation concernant la "forme", notamment du nombre moyen de syllabes par mot :

(1) IS 2,239 IA 2,215 T 2,123 DP 2,096 DF 2,075(8)

Pour tout le corpus examiné la moyenne est de 2,153. Or, la différence entre Isocrate et Démosthène ne s'explique pas par l'emploi plus ou moins fréquent de mots appartenant à ces catégories grammaticales particulièrement riches en monosyllabiques que sont les prépositions, négations, conjonctions, particules et articles : tout au contraire ! Alors que ces catégories totalisent 40% des mots du corpus, on trouve les résultats suivants pour les discours respectifs :

(2)	IS	IA	T	DP	DF
pourcentages	40,2	42,9	40,2	37,6	36,7
différence ER - ET	+ 4	+ 132	+ 4	- 34	- 106(9)

Il faut donc vraiment que Démosthène ait recherché les mots pleins brefs, ou tout au moins évité les formes longues.

On constate d'ailleurs la même tendance pour la longueur des phrases. Quand on calcule le nombre théorique de phrases à partir du nombre de

mots par oeuvre et du nombre de phrases du corpus, on obtient

(3) différence ER - ET I -17 T -1 D +18

Quand on calcule le nombre moyen de mots par phrase, on trouve

(4) pour les petits discours IS 28,3 DP 26,8
pour les longs discours IA 25,2 DF 20,4(10)

On s'attendrait donc à des phrases plus complexes chez Isocrate, plus simples chez Démosthène.

Or, pour les verbes, cette catégorie grammaticale qu'on considère souvent comme indice de la complexité des phrases composées (et qui couvre 21,2% des mots du corpus), on obtient

(5) différence ER - ET I -5 T -57 D +62

Remarquons que chez Thucydide, le manque de verbes est compensé par l'abondance des noms, ce qui n'est pas le cas pour Isocrate :

(6) différence ER - ET
substantifs : I -14 T +35 D -21
adjectifs : I -11 T +42 D -31

Et si l'on recourt ici à l'analyse morphologique, trouve-t-on chez Isocrate l'abondance de formes conjuguées à laquelle on s'attend ? Encore une fois, non; au contraire :

(7) différence ER - ET I -72 T -27 D +99

L'emploi de l'infinitif n'explique pas non plus la différence entre Isocrate et Démosthène :

(8) différence ER - ET I +7 T -25 D +18

Il n'y a finalement que les participes dont les nombres d'occurrences vont dans le sens attendu :

(9) différence ER - ET I +60 T -5 D -55

Que nous apprend maintenant l'analyse syntaxique ? Après déduction des verbes principaux, on obtient pour les verbes conjugués subordonnés et coordonnés :

(10) différence ER - ET I -51 T -21 D +72(11)

Donc, plus nettement encore que les chiffres obtenus aux niveaux des catégories grammaticales et de l'analyse morphologique, les résultats de l'analyse syntaxique montrent que ce n'est pas à la multiplication des propositions subordonnées qu'on doit attribuer la longueur des phrases chez Isocrate. Bien au contraire !

Comment expliquer alors la longueur des phrases d'Isocrate ? Par une préférence pour les formes nominales ? Pas exactement : nous avons même constaté un manque relatif pour les substantifs et les adjectifs — voir (6) — et les écarts pour les infinitifs allaient dans le même sens chez Isocrate et Démosthène — voir (8) —; ce n'était vraiment que par rapport aux participes qu'Isocrate manifeste une telle préférence — voir (9) —. C'est seulement au niveau de la syntaxe, à mon avis, qu'on découvre une tendance

pour la col. 60), j'obtiens 23,6% des mots du corpus, et pour les ouvrages respectifs :

(15) pourcentages : IS 22,4 IA 23,5 T 24,6 DP 23,7 DF 23,6

Donc sensiblement les mêmes pourcentages !

Or, quand j'examine comment ce groupe se répartit selon les trois catégories dont il a été composé, je trouve :

(16) pourcentages	IS	IA	T	DP	DF
a. subst. sans art.	13,1	16,8	31,3	24,9	24,9
b. pronoms	27,3	22,7	22,5	38,1	33,2
c. membres de phrases avec art.	59,6	60,5	46,2	37,0	41,9

En chiffres ronds la comparaison d'Isocrate et de Démosthène donne :

a. subst. sans art.	15	25
b. pronoms	25	35
c. membres avec art.	60	40

distribution qui est sensiblement la même dans les deux discours d'Isocrate d'une part, de Démosthène de l'autre. Ce qui semble indiquer qu'on a de nouveau affaire à une caractéristique de style très nette.

Université de Louvain

J. DE BIE
Assistant chargé d'enseignement

NOTES

1. Les recherches, dont cette communication veut esquisser les méthodes et les perspectives, sont réalisées au L.A.S.L.A. et ont pu profiter de l'expérience, des conseils et de l'aide de toute l'équipe du L.A.S.L.A. Je tiens à leur exprimer ici toute ma gratitude.
2. Catégories grammaticales ou parties du discours traditionnelles et analyse grammaticale des formes : cas, nombre, genre, personne, temps, mode, voix.
3. *Revue*, 1966/1, pp. 1-50
4. Dans une conception "hiérarchique" de la structure de la phrase, le "régissant" est le terme supérieur d'une connexion (cf. L. TЭСNIERE, *Eléments de syntaxe structurale*, Paris, 1959, p. 13), c'est-à-dire le mot en rapport avec lequel le "régé" fonctionne dans la phrase.
5. Seule la ponctuation forte (fin de phrase) est indiquée, notamment sur la carte du mot qui la précède.
6. 7. Termes empruntés à A. W. DE GROOT, *Inleiding tot de algemene taalwetenschap*, Groningen, 1962, pp. 66 et 67.
8. Dans l'exposé qui suit, j'emploie les sigles suivants :
 - IS = ISOCRATE XIII : *Contre les Sophistes* (1328 mots)
 - IA = ISOCRATE VII : *Aréopagitique* (4590 mots)
 - I = ISOCRATE XIII + VII (5918 mots)
 - T = THUCYDIDE II, 35-46 : *Oraison funèbre* (1934 mots)

DP = DEMOSTHENE V : *Sur la Paix* (1472 mots)
DF = DEMOSTHENE IV : *Première Philippique* (3306 mots)
D = DEMOSTHENE V + IV (4778 mots)

9. ER = effectif réel; ET = effectif théorique.
On obtient l'effectif théorique d'une catégorie (ou d'un groupe de catégories) dans chaque oeuvre en répartissant l'effectif global de cette catégorie (ou de ce groupe) dans le corpus proportionnellement au nombre total des mots dans chacune des oeuvres.
10. Ici aussi, la moyenne de T (24,5) est très proche de celle du corpus (24,1).
11. On peut constater d'ailleurs que sur 4778 mots de Démosthène, il y a 149 conjonctions de subordination, tandis que sur les 4590 mots de l'Aréopagitique d'Isocrate, on n'en compte que 89.