

LA LONGUEUR DES MOTS COMME MESURE D'HOMOGENEITE CHEZ HIPPOCRATE*

GILLES MALONEY, WINNIE FROHN VILLENEUVE, ISABELLE DIAZ
(GRAPHISTE)

LES PAGES qui suivent risquent d'être jugées un peu trop nombreuses, surtout lorsqu'elles sont consacrées à des discussions de détail. On doit noter immédiatement que plusieurs théories qui les sous-tendent ne sont pas exposées, car elles ont été soutenues dans deux articles précédents: l'un sur "La longueur des mots dans des oeuvres de prose grecque",¹ l'autre sur "Mots longs et mots courts en grec".² Comme nous transportons le vaste problème de l'homogénéité du *Corpus* hippocratique dans cette petite province de la stylométrie qu'est la longueur des mots, le lecteur qui ne trouve pas de délectation particulière dans des données quantitatives désirera peut-être ne connaître le contenu que de certaines parties de notre travail.

Nous avons donc réparti cet article en cinq sections qu'on pourra repérer aisément par leur numérotation:

1. Présentation de la question—Données de base
2. Trois aspects de la longueur moyenne des mots
3. Détails sur des mots de longueurs précises
4. Aspects de la courbe de distribution des longueurs de mots
5. Conclusions sur l'homogénéité du *Corpus*.

1.1 PRÉSENTATION

La longueur des mots comme mesure d'homogénéité a depuis longtemps fait l'objet d'études. En 1851 Auguste de Morgan écrit dans une lettre que la moyenne des lettres par mot pourrait être un critère de distinction entre deux auteurs.³ Il donne comme exemples les épîtres de St-Paul, ainsi que les œuvres d'Hérodote et de Thucydide. Inspiré par cette lettre,

*Cet article n'aurait sans doute jamais été écrit si le Projet Hippo n'avait pas joui de l'appui généreux du Conseil des Arts du Canada et de l'aide du Ministère de l'Education du Québec.

¹Voir *Cahier des études anciennes* 5 (1976) 59-90.

²Voir *Revue du LASLA* 3 (1975) 1-34.

³Sophia de Morgan, *Memoir of Augustus de Morgan by his wife Sophia Elisabeth de Morgan, with selections from his letters* (London 1882) 215-216 (cité par C. B. Williams, *Style and Vocabulary: Numerical Studies* [London 1970] 4-5, 156). Pour les études sur la longueur des mots en général, on ne peut faire mieux que consulter l'excellente thèse de R. S. Wachal, *Linguistic Evidence, Statistical Inference and Disputed Authorship* (Diss., University of Wisconsin, 1966).

T. C. Mendenhall écrit un article en 1887 et un autre en 1901, dans lesquels il s'intéresse non seulement aux moyennes, mais aussi à la distribution des fréquences des mots par nombre de lettres.⁴ Il note qu'un écrivain emploie toujours la même distribution de longueurs de ses mots, quel que soit l'auditoire auquel il s'adresse, même si (et ce fut sa découverte principale) plusieurs auteurs peuvent utiliser des distributions similaires, par exemple Shakespeare et Marlowe, ou très différentes, tels Shakespeare et Bacon.

Plus récemment, Brinegar examinait, dans un échantillon de la correspondance de Mark Twain et celle de Q. C. Snodgrass la distribution des longueurs de mots dans les lettres ainsi choisies; dans le genre épistolaire M. Twain a gardé la même distribution des longueurs de ses mots toute sa vie. Brinegar conclut que, cette distribution ne ressemblant pas à celle des mots de Snodgrass, les lettres attribuées à ce dernier n'ont pas été écrites par Twain.⁵

Dans des études différentes S. V. F. Waite et G. Herdan trouvent aussi que les distributions des longueurs des mots sont différentes chez des auteurs écrivant en latin, ou encore en anglais ou en allemand ou en russe.⁶ Et W. Fuchs a mené des études sur cinquante textes allemands pour démontrer que les arrangements des longueurs de mots comptés en syllabes et de la longueur des phrases comptées en mots diffèrent non seulement leurs auteurs mais aussi la qualité du style.⁷

Winter et Kondratov décèlent un phénomène similaire en russe, où la longueur des mots augmente dans les œuvres scientifiques et décroît dans le dialogue et la poésie.⁸ M. Ludvikova remarque la même tendance

⁴T. C. Mendenhall, "The Characteristic Curves of Composition," *Science* 11 (1887) 237-249 (Supplement, March 1887) et "A mechanical solution of a literary problem," *Popular Science Monthly* 60 (1901) 97-105 (cités par C. B. Williams [*supra*, n. 1] 5-8, 36-39, 156).

⁵C. S. Brinegar, "Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship," *Journal of the American Statistical Association* 58 (1963) 85-96.

⁶G. Herdan, *The Advanced Theory of Language as Choice and Chance* (New-York 1966) 89-292; S. V. F. Waite, "Approaches to the analysis of Latin prose, applied to Cato, Sallust and Livy," *Revue* 2 (1970) 91-120.

⁷W. Fuchs a écrit plusieurs articles à ce sujet. On trouvera une bibliographie dans W. Fuchs et J. Lauter, "Mathematische Analyse des literarischen Stils," *Mathematik und Dichtung*, H. Kreuzer et R. Gunzenhäuser éd. (Munich 1965) 107-122. Mentionnons aussi W. Fuchs, "Unterschied des Prosastils von Dichtern und anderen Schriftstellen," *Sprachforum* 1 (1955) 234-244. Lubomir Dolezel trouve une corrélation entre la longueur des phrases et la longueur des mots en tchèque moderne: "A Framework for the Statistical Analysis of Style", *Statistics and Style*, L. Dolezel et R. W. Bailey, éd. (New-York 1969) 19.

⁸W. Winter "Styles as Dialectics," *Statistics and Style* (*supra*, n. 7) 3-9. A. M. Kondratov "Information Theory and Poetics: The Entropy of Russian Speech Rythm," *Statistics and Style* (*supra* n. 7) 114.

dans la langue tchécoslovaque.⁹ L'influence du genre sur la longueur des mots est observée aussi par Waite, qui trouve que les discours rapportés dans l'oeuvre de Caton ont une longueur moyenne des mots plus courts que les parties narratives.¹⁰ Muller remarque que les comédies qu'il étudie ont des mots plus courts que les tragédies, et il suggère qu'en général en français "un vocabulaire technique, savant ou simplement recherché augmente la proportion des mots longs . . ." et qu'une "syntaxe soignée tend à réduire la densité du texte en mots grammaticaux, donc en mots courts . . .".¹¹

Tandis que ces recherches laissent entendre que le genre littéraire a tendance à influencer la longueur des mots, Herdan trouve que deux ouvrages de Tolstoy, différents selon leur genre, ont des distributions de longueurs de mots très similaires.¹² De même, Brainard observe que *Julius Caesar*, une tragédie, et *As You Like It*, une comédie, produisent elles aussi des distributions similaires.¹³

1.2 LES OEUVRES ÉTUDIÉES

Nous voulons pour notre part examiner la distribution de la longueur des mots dans le *Corpus* d'Hippocrate afin de savoir d'abord dans quelle mesure elle peut éclairer la question de l'homogénéité du *Corpus*, et ensuite si elle est influencée par le genre ou le contenu d'un ouvrage. Nous prenons comme hypothèse le fait notoire que le *Corpus hippocraticum*, malgré son titre, ne contient pas uniquement des œuvres d'Hippocrate; bien que tous ces livres traitent de médecine et qu'ils aient été écrits, pour la plupart, au cinquième et au quatrième siècle av. J. C., ils relèvent de plusieurs genres très différents, tels que discours, serment, essai, collection de notes.

Du *Corpus* nous examinerons 45 traités, numérotés de 1 à 47, selon l'édition de Littré (voir l'Annexe I). Ils ne comprennent pas les *Prénotions coaques* (numéro d'ordre 23 chez Littré) ni le livre *Des affections* (no 30), parce que ces traités n'étaient pas enregistrés sur le ruban d'ordinateur qui a servi de support à notre étude. Ces 45 œuvres fournissent 244,502 mots composés de 1,242,099 lettres. Pour obtenir ce total nous définissons un mot comme un ensemble de lettres de l'alphabet grec compris entre

⁹M. Ludvikova, "Quantitative Syllable Analysis of Words in Czech," *Prague Studies in Mathematical Linguistics 3* (U. of Alabama, 1971) 27-34.

¹⁰S. V. F. Waite (*supra*, n. 6).

¹¹Ch. Muller, "La longueur moyenne des mots dans le théâtre classique," *Cahiers de Lexicologie 5* (1964) 31, 35.

¹²G. Herdan (*supra*, n. 6) 291-292.

¹³B. Brainard, *Weighing Evidence in Language and Literature: a Statistical Approach*, Mathematical Expositions no. 19 (Toronto-Buffalo 1974) 248.

TABLEAU 1

Quantité de mots de diverses longueurs chez Hippocrate: détail pour 45 traités

Numéro d'ordre du traité	Mots de "n" lettres									
	1	2	3	4	5	6	7	8	9	10
1	179	972	1144	501	653	559	471	409	208	163
2	139	1224	1739	637	878	787	594	507	392	214
3	182	1075	1029	343	512	455	483	369	228	223
4	200	1103	1200	588	628	670	501	383	301	219
5	179	899	941	444	520	629	542	382	293	210
6	67	714	800	564	720	687	567	491	376	297
7	72	605	709	493	751	781	598	446	377	184
8	158	973	1127	390	562	623	401	248	212	129
9	108	440	342	231	221	226	158	151	113	64
10	340	2233	2259	950	1286	1130	872	630	608	383
11	653	4101	4349	1988	2211	2148	1669	1361	978	696
12	224	1043	778	414	522	501	441	340	241	175
13	279	1194	1085	503	704	828	730	544	443	318
14	5	32	55	26	20	31	19	17	24	12
15	15	52	53	31	29	41	31	24	16	15
16	168	930	886	496	541	584	560	376	263	188
17	146	613	692	432	473	542	429	343	258	163
18	139	800	992	571	714	730	604	397	274	150
19	219	854	878	521	599	511	515	416	300	188
20	287	1430	1735	992	1208	1441	1103	798	549	385
21	96	355	378	232	305	198	226	160	119	72
22	97	367	317	199	235	218	253	216	198	164
24	136	571	567	253	308	263	200	141	158	94

deux blancs, l'apostrophe ayant la valeur d'une lettre: δέ et δ' sont deux mots de deux lettres.

1.3 DONNÉES QUANTITATIVES

Toute notre étude s'appuiera sur les données qui constituent le tableau 1 "Quantité de mots de diverses longueurs chez Hippocrate." Dans ce tableau, auquel on devra sans doute se reporter plusieurs fois, les colonnes donnent le total de mots d'une certaine dimension dans chacun des 45 traités, les rangées donnant le total de mots de chaque dimension dans un certain traité. La première colonne à gauche contient le numéro d'ordre des livres dans l'édition de Littré; le titre du livre concerné figure dans l'Annexe I de cet article. En lisant de haut en bas la colonne 11 on aura le nombre de mots de 11 lettres des traités 1 à 45; en parcourant la rangée 11 on saura combien il y a de mots de toutes les longueurs dans le livre *Des articulations*.

TABLEAU 1 (suite)

Mots de "n" lettres										
11	12	13	14	15	16	17	18	19	20	Total
141	62	34	18	4	8	0	0	0	0	5526
174	99	37	23	8	2	1	0	0	0	7455
126	97	70	15	12	2	3	5	0	7	5236
151	91	43	28	12	11	4	2	2	0	6137
146	74	32	17	7	4	0	2	1	0	5322
172	64	44	30	8	2	4	3	1	1	5612
141	42	32	9	7	2	9	3	0	1	5262
89	47	16	8	1	0	0	0	0	0	4984
39	31	12	4	1	0	0	0	0	0	2141
258	142	89	30	16	11	0	0	0	0	11237
489	298	158	89	40	25	4	0	0	1	21258
111	68	39	16	7	2	1	0	0	0	4923
234	102	63	29	11	4	3	1	0	1	7076
7	1	0	0	0	1	0	0	0	0	250
9	3	0	4	0	1	0	0	0	0	324
110	59	34	13	5	3	2	1	0	0	5219
106	57	27	11	8	6	6	3	0	0	4315
86	49	19	11	4	1	2	0	0	1	5544
108	82	33	26	16	3	4	0	0	0	5273
264	123	56	34	11	5	6	6	1	1	10435
46	33	16	7	4	1	1	0	1	0	2250
114	66	45	14	5	5	2	1	0	0	2516
64	51	18	14	7	1	0	1	0	0	2847

Pour ne pas alourdir la présentation nous ne livrons pas le détail de certains calculs subséquents déjà présentés ailleurs,¹⁴ mais on sera capable, si on le désire, d'extraire de ce tableau tous les renseignements dont nous ferons usage.

Ce tableau est d'ailleurs allégé de 3 mots ayant plus de 20 lettres (plus précisément 21, 23 et 25 lettres), qui rendaient la présentation difficile et apportaient peu à la discussion.

2.1 LONGUEUR DU TEXTE ET LONGUEUR DES MOTS

Comme il existe des différences importantes dans l'ampleur des textes, qui vont de 250 mots dans le *Serment* (no 14) à 21,258 mots dans les *Articulations* (no 11), avant de comparer la distribution des longueurs de mots dans les différents traités, nous tenterons de voir s'il existe une relation entre la longueur du texte et la longueur moyenne des mots. En

¹⁴Voir *supra*, notes 1 et 2.

TABLEAU 1 (suite)

Quantité de mots de diverses longueurs chez Hippocrate: détail pour 45 traités

Numéro d'ordre du traité	Mots de "n" lettres									
	1	2	3	4	5	6	7	8	9	10
25	86	687	877	347	416	426	271	240	137	119
26	33	234	342	115	121	179	111	99	73	49
27	83	483	618	239	331	210	233	179	165	91
28	46	235	252	100	181	212	141	103	81	60
29	322	1437	1727	624	724	872	701	449	314	198
31	290	1588	1727	933	819	893	718	464	406	234
32	111	817	1073	402	537	515	353	266	280	148
33	90	582	591	311	459	405	294	206	155	132
34	21	159	162	63	102	113	85	61	45	46
35	43	248	289	124	200	198	140	111	79	43
36	117	1143	1347	552	761	618	503	497	334	220
37	112	1116	1447	426	736	686	507	461	364	196
38	80	845	946	379	504	525	386	347	306	172
39	69	396	452	182	222	272	211	207	140	82
40	317	2076	2599	1219	1372	1724	1238	777	573	359
41	131	869	1022	434	539	640	502	329	253	142
42	480	2563	3115	1381	1865	2479	1754	960	611	413
43	194	1724	2096	961	1533	1832	1243	830	581	401
44	38	314	358	143	162	181	146	109	75	56
45	25	131	164	79	73	63	76	48	56	21
46	50	359	427	199	252	225	165	111	57	38
47	194	1363	1744	689	777	872	639	391	276	140
<i>Corpus:</i>	7020	41949	47430	21701	27287	28823	22384	16394	12290	8066
	1	2	3	4	5	6	7	8	9	10

procédant à cette comparaison nous gardons à l'esprit que la longueur du texte exerce au moins une influence sur le vocabulaire, comme G. Herdan et G. U. Yule l'ont montré:¹⁵ le taux d'apparition de mots nouveaux dans un texte s'abaisse en fonction de l'augmentation de la longueur du texte.

Nous avons établi un relevé comparatif de la longueur des mots et de l'ampleur de chacun des 45 traités; il constitue le tableau 2, où les livres du *Corpus* sont classés selon l'ordre décroissant de la longueur moyenne de leurs mots. Cette compilation ne fait apparaître aucune relation frappante entre les deux variables.

¹⁵Voir par exemple G. Herdan, "Chaucer's Authorship of 'The Equatorie of the Planets'", *Language* 32 (1956) 254-259, pour l'influence de la longueur du texte sur le vocabulaire. Voir aussi G. U. Yule, *The Statistical Study of Literary Vocabulary* (Cambridge 1944) 14 et *passim*.

TABLEAU 1 (suite)

Mots de "n" lettres										
11	12	13	14	15	16	17	18	19	20	Total
72	44	16	6	3	2	0	0	0	0	3749
48	26	13	7	3	3	0	0	0	0	1456
49	29	15	5	1	1	0	1	0	0	2833
31	13	7	7	5	0	0	0	0	0	1474
141	93	61	36	6	10	1	0	0	0	7716
185	110	54	20	19	4	1	1	0	2	8468
108	62	22	15	2	0	0	0	0	0	4711
67	40	11	1	2	2	0	0	0	0	3348
21	10	2	4	2	0	0	1	0	0	897
32	28	10	2	1	0	0	0	0	0	1548
170	89	49	18	4	1	0	0	0	0	6423
139	96	32	16	9	7	0	0	0	0	6350
137	81	33	15	9	2	2	0	0	1	4770
53	34	5	6	1	0	0	0	0	0	2332
189	101	51	35	23	2	1	4	1	0	12661
80	43	30	13	7	1	0	1	0	1	5037
266	121	66	23	7	4	0	0	1	0	16109
284	65	26	21	4	0	0	0	0	0	11795
36	29	7	9	4	0	1	1	0	0	1669
20	13	4	1	5	0	0	1	0	1	781
38	9	5	2	4	0	0	0	0	0	1941
115	49	29	8	2	1	0	0	0	0	7289
5466	2926	1465	720	317	140	58	38	8	18	244,499
11	12	13	14	15	16	17	18	19	20	Total

Nous avons malgré tout mesuré, à l'aide de l'indice de Pearson "r", la corrélation existant entre le nombre de mots par traité (variable X) et la longueur moyenne des mots dans chaque traité (variable Y). Selon cet indice, qui varie entre les valeurs 1 et -1, la corrélation va d'un parallélisme à une inversion parfaite entre X et Y; s'il n'y a pas de corrélation entre X et Y, alors $r = 0$. Or la corrélation entre nos deux variables se fixe à -0.1681 pour les résultats absolus, et à -0.1839 si on utilise le logarithme du nombre de mots par traité.¹⁶

¹⁶La formule de corrélation employée est:

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}$$

Pour une discussion sur la valeur de cet indice voir M. Ezekiel et K. A. Fox, *Methods of Correlation and Regression Analysis*⁸ (New-York 1959) 118-133, ou L.-T. Dayhaw, *Manuel de Statistique* (Ottawa 1966) 118-146.

TABLEAU 2

Longueur moyenne des mots dans 45 traités d'Hippocrate

Numéro d'ordre du traité dans Littré	Longueur moyenne des mots	Nombre de mots dans le traité	Numéro d'ordre du traité dans Littré	Longueur moyenne des mots	Nombre de mots dans le traité
22	5.968 lettres	2 516	37	5.075	6 350
6	5.718	5 612	33	5.034	3 348
7	5.658	5 262	44	5.029	1 669
13	5.426	7 076	2	5.019	7 455
17	5.381	4 315	41	4.989	5 037
20	5.365	10,435	12	4.990	4 923
14	5.336	250	1	4.969	5 526
38	5.259	4 770	40	4.968	12,661
28	5.258	1 474	32	4.968	4 711
15	5.256	324	42	4.967	16,109
5	5.244	5 322	11	4.962	21,258
34	5.233	897	27	4.953	2 833
19	5.223	5 273	10	4.946	11,237
43	5.223	11,795	31	4.880	8 468
26	5.207	1 456	24	4.877	2 847
39	5.169	2 332	9	4.859	2 141
45	5.151	781	29	4.829	7 716
35	5.142	1 548	25	4.813	3 749
21	5.135	2 250	8	4.707	4 984
3	5.130	5 236	46	4.673	1 941
16	5.129	5 219	47	4.651	7 289
18	5.120	5 544			
36	5.102	6 423			
4	5.082	6 137			
			Corpus	5.080	244,502

Il n'existe donc pas de corrélation appréciable entre les deux variables: à peine une tendance à un rapport inverse, selon lequel la longueur moyenne des mots d'un texte tendrait à être inversement proportionnelle à la longueur de ce texte. Comme la corrélation semble pour ainsi dire inexistant, nous retenons qu'il n'est pas absolument nécessaire, pour les fins de notre étude, de tenir compte de la longueur des livres d'Hippocrate. Cette conclusion ressemble à celle où G. Herdan aboutissait déjà.¹⁷

Ces énoncés n'impliquent pas toutefois que la longueur d'un texte n'influence pas la longueur des mots: nous retenons l'hypothèse que la longueur d'un texte n'influence pas de façon significative la longueur moyenne de l'ensemble des mots du texte. Il reste possible que la croissance d'un texte en longueur fasse croître ou décroître la longueur moyenne cumulative des mots. Il est possible qu'au cinq-centième mot d'un texte

¹⁷G. Herdan (*supra*, n. 6) 292: pour la longueur des mots "The difference in sample size appears to make no appreciable difference provided N is large."

gardé dans son ordre original la longueur moyenne soit de 5.555 lettres, mais qu'au millième mot elle soit de 5.455 lettres, et de 5.355 au mille cinq-centième mot.

2.2 LA LONGUEUR MOYENNE DES MOTS

Dans l'ensemble du *Corpus* que nous avons constitué la longueur moyenne des mots s'établit de 5.08 lettres. Nous notons, à titre de curiosité que le traité qui se situe le plus près de la moyenne est celui du *Régime dans les maladies aiguës* (no 4), avec 5.082 lettres.

Pour le moment, sans avoir recours à des techniques plus sophistiquées que le premier coup d'oeil, notre attention est attirée par 6 livres qui paraissent se détacher du groupe. Trois d'entre eux produisent les moyennes les plus élevées: les *Prorrhétiques I* (no 22), ayant 5.968 lettres; les *Epidémies I* (no 6), ayant 5.718 lettres et les *Epidémies III* (no 7), ayant 5.658 lettres. Trois autres ont les moyennes les plus basses: *Les plaies de la tête* (no 8) avec 4.707 lettres, *De la génération* (no 46) avec 4.673 lettres et *De la nature de l'enfant* (no 47), avec 4.651 lettres.

Sans accorder une valeur d'étalon à ces données en tant que critères d'homogénéité, nous sommes assez satisfaits de constater que nous retrouvons isolés de la masse deux paires dont on admet généralement qu'elles appartiennent chacune à un même auteur: *Epidémies I et III* d'une part, *De la génération* et *De la nature de l'enfant* d'autre part.

Nous avons avancé, au début de cette étude, que la longueur moyenne des mots, seule ou en même temps que d'autres mesures, pouvait servir d'indice d'authenticité d'un texte. Nous voulons le mettre à l'épreuve d'abord en tenant compte du groupement des textes selon l'école médicale. Si on définit trois classes d'ouvrages à savoir ceux d'Hippocrate ou de son école, ceux de "l'école" de Cnide, et ceux de médecins philosophes ou théoriciens, nous obtenons la compilation reportée plus bas, qui s'inspire de l'ensemble des opinions de Littré dans ses diverses notices, de L. Bourgey,¹⁸ de R. Joly¹⁹ et de D. D. Lipourlis.²⁰ Les trois groupes

¹⁸Passim dans *Observation et expérience chez les médecins de la collection hippocratique* (Paris 1953) et dans "La médecine grecque des origines à la fin de l'époque classique," *Histoire générale des sciences* 1 (Paris 1957) 276-299.

¹⁹Passim, surtout dans *Hippocrate, Médecine grecque* (Paris, coll. *Idées*, 1964) et dans *Le niveau de la science hippocratique* (Paris, 1966) et "Hippocrates of Cos", *Dictionary of Scientific Biography* (New-York 1972) 418-431. On consultera aussi les opinions du Prof. Joly sur des traités pris individuellement, par exemple, dans *Recherches sur le traité pseudo-hippocratique Du régime* (Paris-Liège 1961) ainsi que ses éditions de 9 traités du *Corpus* dans *Hippocrate* (Paris, tome VI [1ère partie], 1972, tome VI [2e partie], tome XI, 1970).

²⁰Voir dans son excellent ouvrage 'Η παραγωγική κατάληξη -ικος στήν προσωκρατική φιλοσοφία καὶ στὸ ἴπποκρατικὸν Corpus (Thessalonique 1968) 57-78.

sont formés des traités 1 à 26 et 28, 29, 32 à 35 pour les œuvres coaques; des traités 31, 40, 41, 42, 43, 46 et 47 pour les Cnidiens; et 27, 36 à 39, 44, 45 pour les œuvres théoriques. Nous compilons la moyenne générale de chacun des 3 groupes, ainsi que la valeur de la moyenne la plus élevée et celle de la moyenne la plus basse qu'on peut trouver dans chaque groupe:

	Longueur moyenne des mots (1)	Longueur moyenne maximale (2)	Longueur moyenne minimale (3)	Etendue (2)-(3)
Oeuvres coaques	5.12 lettres	5.968	4.707	1.261
Oeuvres cnidiennes	4.96	5.233	4.651	0.582
Oeuvres théoriques	5.11	5.258	4.877	0.381

Les traités cnidiens affichent une moyenne sensiblement plus basse que celle des deux autres groupes, tandis que les œuvres de "philosophes" se situent dans un étendue très restreinte, alors qu'au contraire les livres coaques jouent sur un écart relativement grand. On serait porté à croire en conséquence que la longueur moyenne des mots reflète l'origine commune de certains types d'œuvres; cela n'est pas exclu. Mais l'étendue plus étroite des œuvres cnidiennes et "théoriques" n'est peut-être due qu'à la quantité plus restreinte de livres formant ces groupes: il faudrait classer tout le *Corpus* avant de pousser plus loin l'analyse de cette question.

2.3 LA LONGUEUR MOYENNE ET LE GENRE LITTÉRAIRE

Comme nous l'avons indiqué plus haut, certains pensent que le genre littéraire est susceptible d'influencer la longueur moyenne des mots. Une façon courante de poser la question consiste à se demander si la longueur moyenne des mots est plus élevée dans un ouvrage technique que dans un récit fictif, par exemple. Profitant de la grande diversité qui marque les œuvres du *Corpus*, nous les avons répartis en 8 groupes définis comme suit:

- 1 Un discours: *Des vents* (no 27), généralement accepté comme tel.
- 2 Deux textes courts: *Le serment* (no 14) et *La loi* (no 15), qui constituent des déclarations de principes.
- 3 Des essais, philosophiques ou littéraires: *Ancienne médecine* (no 1), *De l'art* (no 24), *Nature de l'homme* (no 25), *Du régime I* (no 36). Chaque ouvrage comprend introduction et conclusion, une argumentation suivie et bien organisée, un style soigné et traite d'une question générale.

TABLEAU 3

Classement des longueurs de mots selon les genres littéraires

Groupé	Nombre total de mots	Longueur moyenne des mots
1	2 833	4.953 lettres
2	574	5.290
3	92,125	4.967
4	29,112	4.845
5	50,468	5.114
6	48,195	4.950
7	49,123	5.334
8	9 592	5.568

- 4 Essais médicaux: *Airs, eaux et lieux* (no 2), *Maladies I* (no 29), *Maladie sacrée* (no 32), *Génération* (no 46), *Nature de l'enfant* (no 47). Portant sur une question spécifiquement de médecine, ils sont descriptifs, détaillés, concrets, mais non sans prétentions littéraires.
- 5 Traités médicaux: *Pronostic* (no 3), *Régime dans les maladies aiguës* (no 4), *Régime salutaire* (no 26), *Usage des liquides* (no 31), *Régime II, III, IV* (nos 37, 38, 39), *Nature de la femme* (no 43), *Foetus de 7 mois* (no 44), *Foetus de 8 mois* (no 45). Textes bien rédigés, mais sans prétentions littéraires; portent sur des problèmes précis.
- 6 Traités techniques: *Plaies de la tête* (no 8), *Fractures* (no 10), *Articulations* (no 11), *Mochlique* (no 12), *Plaies* (no 33), *Hémorroïdes* (no 34), *Fistules* (no 35). Exposés techniques, ordonnés, détaillés, sur des sujets spécialisés.
- 7 Notes et études: *Appendice aux maladies aiguës* (no 5), *Epidémies I, III* (no 6, 7), *Officine du médecin* (no 9), *Epidémies II à VII* (nos 16 à 20), *Humeurs* (no 21), *Maladies II, III* (nos 40, 41), *Affections internes* (no 42). Description ou étude de cas ou de maladies particulières. Notes et prescriptions sur divers sujets.
- 8 Pensées: *Aphorismes* (no 13), *Prorrhétiques I* (no 22) Collections d'aphorismes et de notes certainement pris dans plusieurs auteurs. Cette répartition, impressioniste et imparfaite, est aussi objective que nous pouvions la faire, si nous presupposons que chaque traité est homogène. Correspond-elle en tout cas à une distribution particulière de la longueur des mots?
- Si on ne tient pas compte de la classe 2, qui risque de fausser l'interprétation à cause de sa taille petite en nombre de mots, on a l'impression qu'un mouvement général se dessine: plus un ouvrage est précis et unifié, plus la longueur moyenne de ses mots est basse; plus il est morcelé

et divers, plus cette moyenne s'élève. En classant les groupes selon la longueur moyenne décroissante des mots on voit apparaître les "pensées," puis les "notes et études," les "traités médicaux," les "essais philosophiques ou littéraires," le discours, les "traités techniques," et enfin les "essais médicaux." La distinction ne paraît donc pas s'établir entre les traités scientifiques et les ouvrages plutôt littéraires. Cette hypothèse expliquerait aussi la moyenne élevée du *Serment* et de *La loi*, qui, malgré leur brièveté, juxtaposent plusieurs idées diverses.

Si une nouvelle classification était établie à partir de tous les traités du *Corpus* à la lumière du principe, il ne serait pas impossible que la variation de la longueur des mots se conforme mieux au degré de précision et d'unité de chacune des œuvres.

3.1 LES MOTS TRÈS LONGS

L'expression "mot très long" est susceptible de prendre la valeur que chacun décide de lui attribuer. Pour notre part nous définissons, sans le justifier autrement, qu'un tel mot est fait de 15 lettres ou plus.

Nous découvrons que, par un hasard qui favorise notre choix, tous les traités ont des distributions qui vont de 1 à 15 lettres au moins. Mais plusieurs n'ont pas de mots mesurant plus de 15 lettres: les livres sur *Les plaies de la tête* (no 8), *L'officine du médecin* (no 9), *L'usage des liquides* (no 28), *La maladie sacrée* (no 32), *Les fistules* (no 35), *Les songes* (no 39), *La nature de la femme* (no 43), *La génération* (no 46). En comparant néanmoins l'importance accordée aux mots très longs, c'est-à-dire ceux de 15 lettres ou plus, relativement à l'ensemble du vocabulaire, on s'aperçoit qu'elle ne dépasse jamais 1%. En groupant les livres sous cet aspect on obtient le classement suivant:

TABLEAU 4
Importance des mots de 15 lettres ou plus chez Hippocrate

Pourcentage des occurrences	Numéro des traités selon Littré	Total de traités
0 à 0.10%	8—9—32—35—36—39—42—43—47	9
0.11 à 0.20	2—12—18—25—27—33—41	7
0.21 à 0.30	1—5—10—13—16—20—29—37—38—40—46	11
0.31 à 0.40	6—11—14—15—21—24—28—31—34—44	10
0.41 à 0.50	4—7—19—26	4
0.51 à 0.60	3—17—22	3
0.81 à 0.90	45	1

Un traité se sépare de tous les autres: *Le foetus de 8 mois* (no 45), dont la proportion de mots très longs est, avec 0.896%, nettement supérieure.

Les mots très longs attirent l'attention sur une troisième caractéristique: ils se situent dans une longue glissade qui, partant des mots de 6 ou 7 lettres fait baisser de façon régulière le nombre de mots de chaque catégorie jusqu'à ceux de 20 lettres. De ce point de vue un traité parmi tous les autres fait montre d'un comportement singulier: c'est le *Pronostic* (no 3), où la quantité de mots très longs s'accroît au lieu de baisser. On compte en effet (voir rangée 3 du tableau 1), après 12 mots de 15 lettres, 3 de 17 lettres, 5 de 18 lettres et 7 de 20 lettres. D'ailleurs près de la moitié (7 sur 18) de tous les mots de 20 lettres ou plus recensés dans notre *Corpus* se lisent dans le *Pronostic* (no 3). Le seul autre livre où cette tendance semblerait se dessiner est celui de *Epidémies III* (no 7), avec la série de mots très longs ayant les totaux 7, 2, 9, 3, 0, 1.

3.2 MOTS COURTS, MOYENS ET LONGS

En faisant référence à des expressions souvent utilisées nous répartissons la masse des mots en trois catégories pour mettre à l'épreuve la fécondité de ces divisions: mots courts, mots moyens, mots longs. Puisque nous devons placer des bornes ici encore, nous définissons comme "mot court" celui qui comprend de 1 à 4 lettres, comme "mot long" celui qui prend 10 lettres et plus, comme "mot moyen" celui qui est fait de 5 jusqu'à 9 lettres.²¹

Dans le prochain tableau chaque rangée produit un total de 100%, les chiffres étant arrondis à la première décimale.

En se guidant sur les moyennes générales on extraiera quelques traités qui représentent cette moyenne de façon satisfaisante, tels le *Serment* (no 14) ou *Les plaies* (no 33). Ailleurs, même si la division adoptée ne fournit que 3 éléments à combiner, on voit que des livres seront à égalité dans une catégorie mais présenteront des variations considérables dans les deux autres, tels les *Epidémies III* (no 7) avec une série 18.1-65.8-16.1 par rapport aux *Prorrhétiques I* (no 22) avec une série 17.2-51.7-31.1. Dans l'importance qu'ils accordent à certains types de mots les écrivains se distinguent en faisant varier par exemple les mots longs dans un rapport du simple au triple: elle va de 11.2 dans *La nature de l'enfant* (no 47) jusqu'à 31.1% dans les *Prorrhétiques I* (no 22) et inversement dans ces mêmes livres les mots courts passent du simple au double.

²¹W. Winter considère qu'un mot long en russe est fait de 9 lettres ou plus. Voir *supra*, n. 8. Nous avons justifié de façon plus détaillée cette répartition en 3 classes dans "Mots longs et mots courts en grec," *Revue 3* (1975) 1 suiv.

TABLEAU 5

Espace relatif occupé par chaque catégorie de mots

Numéro d'ordre du traité chez Littré	Mots de 1 à 4 lettres	Mots de 5 à 9 lettres	Mots de 10 lettres et plus	Numéro d'ordre du traité chez Littré	Mots de 1 à 4 lettres	Mots de 5 à 9 lettres	Mots de 10 lettres et plus
1	27.6%	54.9%	17.5%	26	26.2	51.5	22.3
2	27.7	55.7	16.6	27	27.5	57.4	15.1
3	25.3	50.9	23.8	28	21.6	60.9	17.5
4	26.8	52.7	20.5	29	29.2	54.1	16.7
5	23.6	56.8	19.6	31	30.0	52.3	17.7
6	19.2	59.2	21.6	32	28.1	55.1	16.8
7	18.1	65.8	16.1	33	25.3	58.3	16.4
8	30.0	56.5	13.5	34	22.9	57.0	20.1
9	28.3	55.7	16.0	35	23.9	59.9	16.2
10	27.7	53.7	18.6	36	26.4	55.0	18.6
11	28.3	52.5	19.3	37	26.0	56.8	17.2
12	25.6	55.3	19.1	38	24.4	55.4	20.2
13	20.7	57.1	22.2	39	24.4	59.2	16.4
14	25.3	57.8	16.9	40	27.2	59.2	13.6
15	23.6	55.4	21.0	41	26.5	59.4	14.1
16	24.9	57.9	17.2	42	25.6	62.1	12.3
17	22.3	58.9	18.8	43	22.4	63.7	13.9
18	24.7	62.8	12.5	44	27.5	53.2	19.3
19	24.1	57.0	18.9	45	27.2	53.8	19.0
20	22.0	60.2	17.8	46	31.4	56.9	11.7
21	24.8	57.5	17.7	47	32.2	56.6	11.2
22	17.2	51.7	31.1				
24	28.7	50.9	20.4				
25	30.4	53.7	15.9				
				<i>Corpus</i>	25.8%	57.0%	17.2%

Pour le moment nous nous contentons de mettre en lumière les ouvrages qui se distinguent par leurs écarts à la moyenne:

- les *Epidémies I* (no 6) utilisent peu de mots courts et beaucoup de mots longs;
- le traité *De l'art* (no 24) a peu de mots moyens et ne favorise pas l'une des deux autres catégories;
- dans les *Epidémies III* (no 7) on trouvera peu de mots courts et beaucoup de mots moyens;
- les mots courts sont fréquents et les longs sont rares dans les *Plaies de la tête* (no 8), la *Génération* (no 46) et la *Nature de l'enfant* (no 47) et cette double caractéristique se rencontre également, quoique moins marquée, dans *La nature de l'homme* (no 25), *Les lieux dans l'homme* (no 31) et les *Maladies I* (no 29);

- il y a peu de mots moyens et beaucoup de longs dans le *Pronostic* (no 3) et dans le *Régime salutaire* (no 26);
- inversement on lira beaucoup de mots moyens et peu de longs dans les *Epidémies V* (no 18), les *Affections internes* (no 42) et la *Nature de la femme* (no 43);
- enfin les *Prorrhétiques I* (no 22) se tiennent à part, ayant à la fois peu de mots courts, peu de mots moyens et beaucoup de mots longs.

3.3 LES MOTS DE 2 ET 3 LETTRES

Deux autres aspects de la distribution de la longueur des mots vont nous arrêter: les mots de 2 et 3 lettres d'une part, ceux de 5 et 6 lettres de l'autre. Ces quatre classes se groupent deux à deux à la fois par leur voisinage immédiat, par les quantités équivalentes de leurs individus à l'intérieur de chacun des 2 groupes, par leur isolement du reste de la distribution (puisque les catégories de 1, de 4 et de 7 lettres sont nettement décalées par rapport à nos deux groupes et de la sorte les font ressortir), parce qu'ils s'adjudgent les 4 fréquences les plus élevées parmi les 20 classes de longueurs de mots, et qu'ils rassemblent des quantités importantes de mots: 36.6% pour le premier groupe et 23% pour le second dans l'ensemble des traités.

Le total combiné des mots de 2 et de 3 lettres varie d'un livre à l'autre d'autant bas que 25% jusqu'à 42.5%. Pour nous en tenir aux comportements les plus excentriques, examinons à nouveau les extrêmes. Trois livres utilisent en faible proportion les mots de ce groupe: les *Epidémies I* (26.98% no 6), les *Epidémies III* (24.97% no 7) et les *Prorrhétiques I* (27.19% no 22); ceux où la proportion de mots de 2 et 3 lettres est le plus élevée sont les *Plaies de la tête* (42.13% no 8), la *Nature de l'homme* (41.71% no 47).

Que le total des mots de 2 et 3 lettres soit plus ou moins élevé, il peut arriver qu'un écrivain favorise une catégorie en particulier toutefois; cette habitude est marquée dans *Les airs, les eaux et les lieux* (no 2), le *Régime salutaire* (no 26) et dans les *Maladies II* (no 40). A chaque endroit la différence se manifeste en faveur des mots de 3 lettres. Ailleurs il arrive que la quantité des uns et des autres soit très rapprochée; dans les livres des *Fractures* (no 10) et des *Articulations* (no 11), par exemple, dans la *Loi* (no 15), les *Plaies* (no 33), les *Hémorroïdes* (no 34) et surtout dans *l'Art* (no 24).

Dans la plupart des livres on compte plus de mots de 3 que de mots de 2 lettres; dans certains cas la tendance s'inverse, comme dans le *Pronostic* (no 3) et les *Epidémies II* (no 16). Mais la différence en faveur des mots de 2 lettres est la plus nette dans les *Aphorismes* (no 13), les *Prorrhétiques I*

(no 22), l'*Officine du médecin* (no 9) et encore plus dans le *Mochlique* (no 12).

En résumé il nous semble que nous nous attachons à une partie quantitativement importante du vocabulaire; dans plusieurs livres 2 mots sur 5 sont de 2 ou 3 lettres. Ailleurs la fréquence de ces mots est presque deux fois moins grande. Or nous manipulons ici de grands nombres: dans le traité des *Articulations* (no 11) ces deux types de mots s'élèvent à près de 8.500 emplois. Une différence prononcée entre les deux est donc susceptible de correspondre à des quantités non négligeables (plus de 500 mots dans *Les airs, les eaux et les lieux* no 2). Nous avons aussi souligné les endroits où les mots de 2 lettres dépassent ceux de 3 lettres, puisqu'ils s'inscrivent à l'encontre de la tendance générale.

La valeur de ces répartitions en tant qu'instruments de mesure est plus difficile à évaluer quand on relève autant de mots d'une classe que de l'autre. Faut-il s'étonner ou trouver normal que les deux groupes soient à égalité aussi bien lorsque leur total combiné est bas que s'il est élevé? Dans le court texte de la *Loi* (no 15) est-il étonnant de dénombrer 52 mots de 2 lettres et 53 de 3 lettres? Dans le texte beaucoup plus long de l'*Art* (no 24) est-il normal qu'il y ait 571 occurrences d'une sorte et 567 de l'autre? A cette question nous n'offrons pas de réponse et nous négligerons en conséquence les traités où les deux classes de mots arrivent à peu près à égalité.

Il nous paraît donc que les traités qui sortent du rang sont les *Epidémies I* (no 6), les *Epidémies III* (no 7) et les *Prorrhétiques I* (no 22), à cause de la faible proportion de leurs mots de 2 et 3 lettres combinés et de leur proportion de mots de 3 lettres qui est la plus basse de tous les livres. A cause de leur préférence pour la classe de 3 lettres nous plaçons sur une même ligne *Les airs, les eaux et les lieux* (no 2) et le *Régime salutaire* (no 26). Et parce qu'ils favorisent la classe de 2 lettres nous plaçons ensemble le *Mochlique* (no 12) et l'*Officine du médecin* (no 9), tout en notant que la même tendance se manifeste à un moindre degré dans les *Aphorismes* (no 13) et les *Prorrhétiques I* (no 22) et encore moins fortement dans le *Pronostic* (no 3) et les *Epidémies II* (no 16).

3.4 LES MOTS DE 5 ET 6 LETTRES

Le second sommet de la courbe de distribution des longueurs de mots est constitué, pour le *Corpus*, par les classes de 5 et 6 lettres. Ce groupe est moins abondant que celui des 2 et 3 lettres, puisqu'il rassemble environ 23% des occurrences dans les textes. Comme il est naturel, les écarts entre les traités sont moins impressionnantes en nombres absolus mais ils sont à certains endroits bien marqués.

C'est ainsi que le *Pronostic* (no 3), les *Prorrhétiques I* (no 22) et le

traité du *Foetus de huit mois* (no 45) sont les livres où la proportion des mots de 5 et 6 lettres est la plus basse, puisqu'elle s'y tient entre 17 et 18% seulement. A l'opposé, le total grimpe à 29.1% dans les *Epidémies III* (no 7), à 28.5% dans la *Nature de la femme* (no 43) et à 26.9% dans les *Affections internes* (no 42). Le total est également de plus de 25% dans 6 autres livres, dont les *Epidémies I* (no 6).

Pour sa part, la quantité des mots de 6 lettres est légèrement supérieure à celle des mots de 5 lettres; dans quatre traités sa supériorité est nette, soit dans le *Serment* (no 14) et dans la *Loi* (no 15), dans les *Affections internes* (no 42) et dans le *Régime salutaire* (no 26). Dans un seul traité la différence favorise clairement les mots de 5 lettres: celui *Des humeurs* (no 21).

Quant aux variations à l'intérieur des 2 catégories, elles sont les plus sensibles dans le *Serment* (no 14), la *Loi* (no 15) et le *Régime salutaire* (no 26), où la proportion du vocabulaire prise par les mots de 5 lettres n'est même pas de 9%. Elles isolent aussi les *Epidémies III* (no 7), traité où cette proportion est la plus élevée pour tous les livres, puisqu'elle dépasse 14%. Quant aux mots de 6 lettres, c'est *Le foetus de huit mois* (no 45) qui en utilise relativement le moins (à peine 8%), tandis que *Les affections internes* (no 42) et la *Nature de la femme* (no 43) en emploient le plus (15.4 et 15.5% respectivement).

Bref, après cet examen fait à travers la lunette du groupe des 5 et 6 lettres, certains traités attirent l'attention. D'abord, si le *Foetus de Huit mois* (no 45) est faible, cela est dû à sa carence de mots de 6 lettres. Au contraire dans les *Epidémies III* (no 7) le groupe étudié est abondant, à cause surtout du grand nombre de mots de 5 lettres. Les *Affections internes* (no 42) et la *Nature de la femme* (no 43) utilisent beaucoup de mots du groupe, surtout ceux de la classe de 6 lettres. Dans le *Régime salutaire* (no 16), le *Serment* (no 14) et la *Loi* (no 15) la proportion des mots de 5 lettres est basse, surtout lorsqu'on la compare à celle des occurrences de 6 lettres. Le traité *Des humeurs* (no 21) lui, se singularise par la préférence qu'il montre pour les mots de 5 lettres.

4.1 LE MINIMUM ENTRE LES MODES

Le tableau 1 fait voir que, dans le *Corpus* (dernière rangée du tableau) le nombre de mots varie d'une catégorie à l'autre: il est modeste dans la classe de longueur 1, il croît dans les classes 2 et 3, puis s'abaisse pour s'élever à nouveau dans les classes 5 et 6 et décroître constamment dans la suite. Ces deux sommets successivement atteints sont des modes, constitués dans l'ensemble du *Corpus* par les mots de 3 et 6 lettres.

Il est facile de vérifier que pour tous les traités le nombre de mots de 4 lettres est inférieur à celui des mots de 3 lettres (*grossost modo* dans le

rapport du double au simple). Il arrive cependant que le point le plus bas entre les deux sommets se trouve dans les mots de cinq lettres. Cela se produit dans l'*Officine du médecin* (no. 9), le *Serment* (no 14), la *Loi* (no 15) et les *Lieux dans l'homme* (no 31). Il ne semble pas exister non plus de rapport entre ce phénomène et la place de l'un ou l'autre des deux sommets. On doit dire simplement que dans certains ouvrages la fréquence des mots de 4 et 5 lettres est plus basse qu'on s'y attend.

Il reste un livre qui, lui, exagère ce comportement: le *Foetus de huit mois* (no 45), où le creux, situation unique, ne se rencontre que dans la classe de 6 lettres, le second sommet de la courbe étant à 7 lettres. Du même coup la liste des œuvres nous indique que l'autre endroit où le second sommet se trouve à 7 lettres est les *Prorrhétiques I* (no 22). Nous retenons donc le nom de ces deux livres à cause de leur excentricité. Et comme l'*Officine du médecin* (no 9) est le seul ouvrage où à la fois le premier mode est à 2 lettres, le minimum à 5 lettres et le second mode à 6 lettres, nous garderons son nom aussi.

4.2 ETENDUE ET DISPERSION

Une mesure utilisée couramment dans l'analyse de distribution est celle de l'*écart type*, qui est une sorte de synthèse des écarts qui existent entre diverses valeurs formant un groupe et la valeur moyenne de ce groupe. On la décrit souvent par la formule simple

$$\sigma = \sqrt{\frac{\sum (x_i - m)^2}{N}}$$

où σ = écart type
 x = valeur mesurée
 m = moyenne générale
 N = fréquence totale
 \sum = somme

En additionnant ensuite et en soustrayant de la moyenne la valeur trouvée pour σ , on détermine l'étendue qui réalise le meilleur compromis, si l'on peut dire, c'est-à-dire celui grâce auquel on cernera le plus d'individus possible dans les limites les plus rapprochées possible. Dans le prochain tableau nous rassemblons la valeur de $\pm\sigma$ ainsi que les limites suggérées par l'addition et la soustraction de σ à la longueur moyenne (voir tableau 2) des mots dans chaque traité, et ensuite la proportion du total des occurrences qui est contenue dans les limites de la valeur "moyenne $\pm 1\sigma$ ".

4.3 LA ZONE DE FORTE DENSITÉ

Une fois ces limites déterminées, il est possible de calculer aisément la proportion ou le nombre des individus qu'elles contiennent. Ainsi pour

TABLEAU 6

Dispersion de la longueur moyenne des mots dans les limites de $\pm 1\sigma$

Numéro d'ordre du traité selon Littré	Valeur de σ	Etendue de $\pm 1\sigma$		Proportion de mots dans $\pm \sigma$
		Limite inférieure	Limite supérieure	
1	2.803	2.160	7.776	62.6%
2	2.758	2.261	7.777	61.4
3	3.126	2.004	8.256	62.0
4	2.956	2.126	8.038	62.4
5	2.909	2.335	8.153	60.0
6	2.872	2.845	8.590	60.1
7	2.700	2.959	8.358	61.4
8	2.616	2.091	7.323	61.8
9	2.816	2.040	7.673	59.1
10	2.867	2.078	7.813	60.8
11	2.895	2.067	7.857	62.3
12	2.930	2.060	7.920	59.3
13	3.023	2.403	8.449	58.7
14	2.794	2.542	8.130	56.5
15	2.970	2.286	8.226	60.9
16	2.820	2.307	7.949	60.4
17	2.892	2.488	8.273	61.3
18	2.596	2.523	7.716	60.9
19	2.936	2.288	8.159	61.3
20	2.792	2.574	8.157	61.0
21	2.860	2.275	7.995	62.0
22	3.295	2.673	9.263	58.3
24	2.968	1.909	7.845	61.9
25	2.680	2.134	7.493	62.4
26	2.991	2.216	8.198	62.3
27	2.728	2.224	7.681	60.5
28	2.809	2.450	8.067	59.8
29	2.823	2.006	7.652	63.9
31	2.817	2.062	7.697	62.6
32	2.767	2.201	7.735	60.7
33	2.686	2.348	7.721	59.8
34	2.866	2.367	8.099	59.2
35	2.748	2.394	7.890	60.5
36	2.834	2.269	7.936	60.5
37	2.817	2.257	7.892	60.5
38	2.932	2.327	8.191	59.4
39	2.804	2.364	7.973	59.0
40	2.645	2.323	7.613	61.5
41	2.704	2.294	7.643	60.9
42	2.562	2.404	7.529	61.1
43	2.574	2.649	7.797	59.0
44	2.892	2.137	7.921	62.4
45	3.016	2.135	8.167	62.8
46	2.504	2.170	7.177	62.6
47	2.522	2.128	7.173	62.6
<i>Corpus</i>	2.806	2.558	7.886	61.0%

l'ensemble du *Corpus* la limite $\pm\sigma$ englobe 61% de tous les mots; la concentration la moins grande appartient au *Serment* (no 14): 56.55% et la plus dense aux *Maladies I* (no 29): 63.9%, livres qui sortent tous deux du rang par leur supériorité sur leurs plus proches voisins.

Le calcul de σ établit pour l'ensemble du *Corpus* une limite inférieure à 2.558 lettres, c'est-à-dire presque exactement entre les mots de 2 et ceux de 3 lettres. A l'autre extrémité la limite est fixée à 7.886 lettres. Nous ajustons donc ces deux bornes à la réalité pour retenir que la concentration optimum des mots, en ce qui relève de leur longueur, repose dans les occurrences mesurant inclusivement de 3 à 8 lettres. Cette zone contient 163,734 mots, soit 67.1% de l'ensemble.

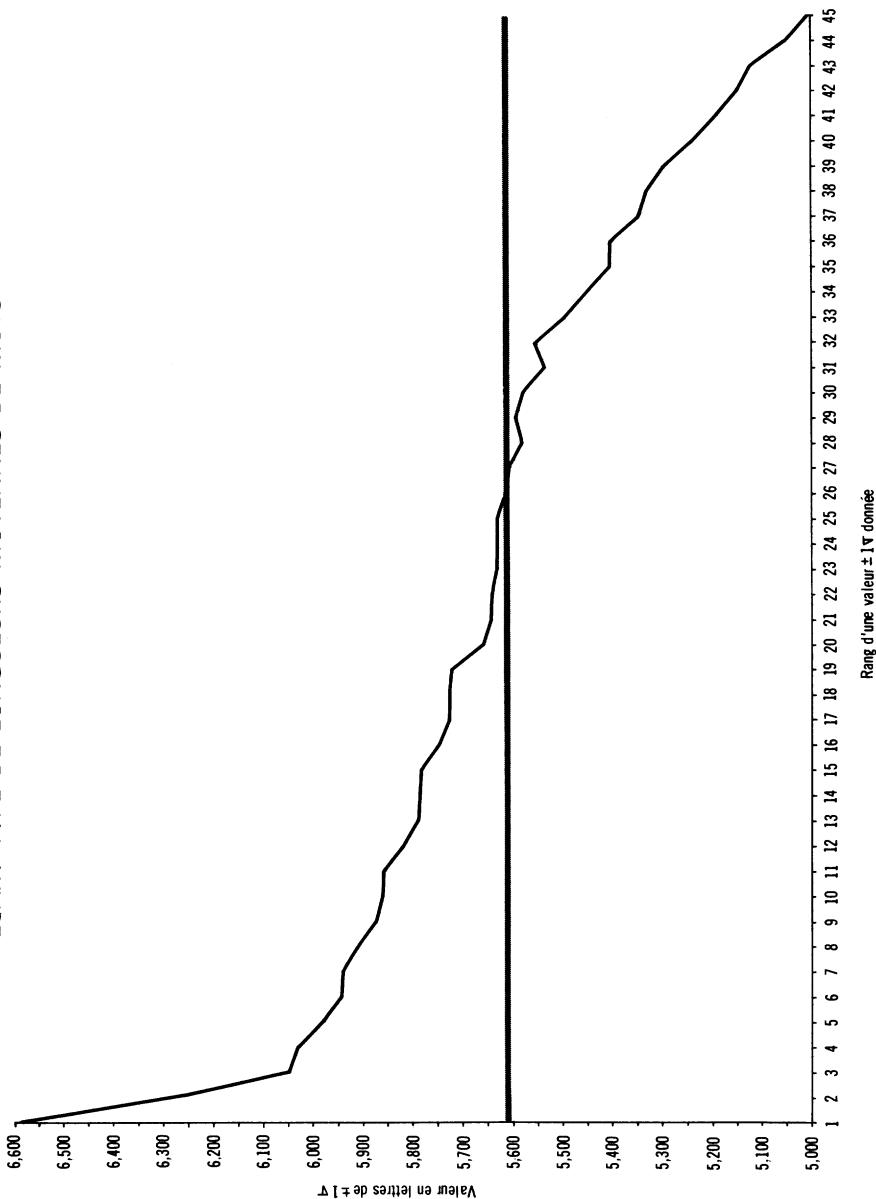
L'examen des résultats qu'on obtient, en ajoutant ou en soustrayant la valeur du σ à celle de la longueur moyenne des mots d'un livre, montre que le produit minimum, c'est-à-dire les valeurs de la limite inférieure (voir tableau 6), se situe dans une zone très étroite qui va de 2.004 pour le *Pronostic* (no 3) à 2.959 pour les *Epidémies III* (no 7). Une seule exception: le traité de l'*Art* (no 24), qui affiche une valeur de 1.909. La zone de la limite supérieure est plus étendue, ce qui est normal dans une courbe dissymétrique; elle se tient dans la région de 7 et de 8 lettres, avec une seule exception, elle aussi: les *Prorrhétiques I* (no 22), où la limite supérieure est portée à 9.263. Ces deux ouvrages, l'*Art* (no 24) et les *Prorrhétiques I* (no 22), se singularisent par ce point. Il faudrait noter également que les *Epidémies I* (no 6) ont une limite qui se situe parmi les plus élevées, près de celle des *Epidémies III* (no 7).

On obtiendra une vue analogue, mais non identique, du classement relatif des traités en les rangeant selon la dimension de la zone délimitée par $m \pm \sigma$. Cette dimension s'établit par le tableau 6, soit en doublant la valeur de σ soit en soustrayant le chiffre représentant la limite inférieure de celui de la limite supérieure. Nous illustrerons le résultat d'un tel classement dans le schéma 7, qui transcrit en ordonnée la valeur de 2 sigmas, tandis qu'en abscisse il représente le rang (allant de 1 à 45) occupé par cette valeur lorsqu'on classe les traités en ordre décroissant de ce point de vue.

Il apparaît par le diagramme (page 115) que l'ensemble des livres se range selon une courbe passablement régulière, qui se distribue en deux sections presque égales de chaque part de la valeur moyenne 5.61. Il nous semble qu'un livre se sépare du reste à l'extrême gauche de la courbe: les *Prorrhétiques I* (no 22). Les autres traités qui se signalent le plus par leur écart type élevé, et par conséquent à cause de la grande dispersion de leurs mots autour de la moyenne sont, dans l'ordre: le *Pronostic* (no 3) et les *Aphorismes* (no 13). A l'autre extrémité, ceux dont la zone est la moins vaste, et par conséquent dont les mots se groupent avec le plus de densité autour de la moyenne, sont les *Affections internes*

TABLEAU 7

ÉCART TYPE DE LONGUEURS MOYENNES DE MOTS



(no 42), la *Nature de la femme* (no 43), la *Génération* (no 46) et la *Nature de l'enfant* (no 47), dernier élément de la courbe.

5.1 RÉSUMÉ

Conclusions particulières sur l'homogénéité du *Corpus*.

Dans les pages précédentes 34 des 45 traités qui formaient notre *Corpus* se sont distingués à un moment ou à l'autre: on les trouvera regroupés plus loin dans le tableau 8. Certains titres sont apparus plusieurs fois, d'autres à une seule occasion mais avec une netteté particulière.

Quelques livres ont présenté une seule caractéristique notable: les *Epidémies II* (no 16), par exemple, ont plus de mots de 2 que de 3 lettres, mais le contraire vaut pour les *Maladies II* (no 40). Dans les *Humeurs* (no 21) il y a plus de mots de 5 que de 6 lettres, et les *Maladies I* (no 29) montrent à la fois une des plus fortes proportions de mots de 2 et 3 lettres et le plus grand pourcentage de leur population dans les limites de $\pm 1\sigma$. Il serait étonnant que ces 4 ouvrages soient parents.

D'autres traités au contraire se sont rapprochés, quoique avec une force d'évidence variable. Ainsi les *Fractures* (no 10) et les *Articulations* (no 11) se rejoignent en ce qu'ils utilisent à peu près autant de mots de 2 que de 3 lettres; mais il en va de même aussi dans les *Plaies* (no 33) et les *Hémorroïdes* (no 34).

Les ressemblances sont plus nettes entre le Premier et le Troisième livre des *Epidémies* (nos 6 et 7), encore que le Troisième livre ait plus de caractères prononcés que l'autre: beaucoup de mots de 5 et 6 lettres et production soutenue de mots très longs.

Les pages sur la *Nature de la femme* (no 43) et les *Affections internes* (no 46), pour leur part, ont en commun beaucoup de mots moyens et peu de longs, beaucoup de mots de 5 et encore plus de 6 lettres, leur population est dense autour de la moyenne, et tous deux ils emploient très peu de mots très longs.

On aura remarqué depuis le début de cette étude l'apparition souvent simultanée de la *Génération* et de la *Nature de l'enfant* (nos 46 et 47), qui ont pour traits communs beaucoup de mots courts, et en particulier beaucoup de mots de 2 et 3 lettres, une valeur très basse du σ , peu de mots longs et une longueur moyenne des mots parmi les plus basses. Ces deux traités fournissent à coup sûr l'exemple le plus net de ressemblance dans tout le *Corpus*.

A la suite des traités isolés, puis des traités rapprochés, nous suggérons une troisième section possible. Nous sommes mis sur une piste en découvrant qu'une faible proportion de mots longs, une forte proportion de mots courts, et en particulier un grand nombre de 2 et de 3 lettres carac-

térisent à la fois la *Nature de l'homme* (no 25) et les *Lieux dans l'homme* (no 31). Or ces mêmes caractéristiques rapprochent aussi la *Génération* (no 46) et la *Nature de l'enfant* (no 47). Mais un autre ouvrage semble bien plus près de ces deux derniers: les *Plaies de la tête* (no 8), qui, aux caractéristiques communes à ces quatre traités ajoute une moyenne basse de la longueur générale des mots qui se trouve aussi dans la *Génération* (no 46) et la *Nature de l'enfant* (no 47). Or si on n'oublie pas le duo *Affections internes*—*Nature de la femme* (nos 42 et 43), on lui découvre des sympathies pour celui de la *Génération*—*Nature de l'enfant* (nos 46 et 47). Ces sept traités constituerait un groupement hypothétique.

Dans une quatrième section nous placerions des livres formant de petits groupes dont la cohérence nous semble douteuse. Les traités des *Plaies* (no 33) et des *Hémorroïdes* (no 34), avons-nous dit, ont tous deux une proportion presque identique de mots de 2 et de 3 lettres: ils se séparent pourtant sur bien d'autres points.

Les deux courts textes du *Serment* (no 14) et de la *Loi* (no 15) ont tous deux plus de mots de 6 que de 5 lettres, peu de mots de 5 lettres, et un creux dans la courbe de distribution des longueurs précisément dans cette cellule de 5 lettres. D'autre part dans l'*Art* (no 24) la quantité de mots de 2 lettres et de 3 lettres est égale, comme dans la *Loi* (no 15). Et la proportion de mots très longs dans ces trois ouvrages est très ressemblante de l'un à l'autre (voir tableau 4).

A une occasion le *Mochlique* (no 12) s'est rencontré avec l'*Officine du médecin* (no 9): tous deux emploient plus de mots de 2 lettres que de 3 lettres. L'*Officine* s'était cependant signalée à plusieurs endroits de notre étude, et se rapprochait des *Lieux dans l'homme* (no 31): c'est dans la classe des mots de 5 lettres que pour tous deux se trouve le minimum entre les modes dans la courbe de distribution des longueurs de mots, et la valeur de leur σ est presque identique: 2.816 pour l'*Officine* et 2.811 pour les *Lieux dans l'homme*.

Nous avons vu également que les *Epidémies III* (no 7) était le seul traité, à part le *Pronostic* (no 4) et avec moins de netteté, où l'on décèle une tendance à l'accroissement du nombre des mots très longs. Les deux ouvrages comprennent aussi beaucoup de mots de longueur moyenne. Même si le *Pronostic* (no 4) ne ressemble pas aux *Epidémies I* (no 6), comme nous avons rapproché les *Epidémies I* et *III* (nos 6 et 7), nous leur ajouterons le *Pronostic*.

Un traité d'humeur vagabonde fut le *Régime salutaire* (no 26), qui s'est dirigé tantôt du côté du *Pronostic*, tantôt de celui des *Airs, des eaux et des lieux* (no 2), tantôt vers les *Humeurs* (no 21) ou les *Maladies II* (no 40) ou vers les *Epidémies V* (no 18), tantôt du côté du *Serment* (no 14) et de la *Loi* (no 15). Mais, puisque l'ami de tous n'est l'ami de personne, nous enverrons le *Régime salutaire* (no 26) dans un nouveau secteur, celui des

TABLEAU 8

Traités rapprochés	Groupe hypothétique	Parentés possibles
1-Epidémies I Epidémies III	Plaies de la tête Nature de l'homme	1-Loi Art Serment
2-Génération Nature de l'enfant	Nature de l'enfant Génération Affections internes	2-Mochlique Officine du médecin
3-Fractures Articulations	Nature de la femme	3-Officine du médecin Lieux dans l'homme
4-Affections internes Nature de la femme	<i>Peu caractérisées</i> 1-Des liquides 2-Maladie sacrée	4-Epidémies I et III Pronostic
<i>Livres isolés</i>	3-Fistules 4-Songes 5-Régime (mal. aiguës) 6-Epidémies V 7-Airs, eaux et lieux	5-Plaies Hémorroïdes
1-Prorrhétiques I 2-Foetus de 8 mois 3-Régime salutaire	<i>Exclusions mutuelles</i> 1-Epidémies II 2-Maladies I 3-Maladies II 4-Humeurs	6-Prorrhétiques I Aphorismes

traités isolés. Il y sera rejoint par le *Foetus de huit mois* (no 45), qui s'est distingué par la très basse proportion de ses mots de 6 lettres et surtout par la proportion élevée des mots longs de 15 lettres et plus.

Reste enfin le grand excentrique qui s'est à chaque fois signalé à l'attention: les *Prorrhétiques I* (no 22). Chez lui se combinent peu de mots courts et peu de moyens à beaucoup de longs; il est le seul dont la courbe ait des sommets à 2 puis à 7 lettres; il emploie au total très peu de mots de 2 et 3 lettres; il produit la valeur la plus élevée de sigma, et la longueur moyenne de ses mots, comme on s'en doute, figure parmi les plus hautes. Aucun livre n'est à l'abord aussi fortement caractérisé.

Le dernier traité que nous retenons est celui des *Aphorismes* (no 13), à cause de sa petite quantité de mots courts et sa grande quantité de mots longs, ainsi que pour la valeur élevée de son écart type. Ces trois caractéristiques très nettes se rencontrent seulement dans les *Prorrhétiques I* (no 22). Si l'on devait rapprocher un autre ouvrage des *Prorrhétiques I*, il faudrait choisir le *Pronostic* (no 4), à cause de ses mots moyens et longs et de la valeur de son écart type.

En schématisant ces résultats nous pouvons grouper les œuvres en

cinq lots, selon que des traités paraissent offrir des analogies ou ne pas présenter de comportements quantitativement équivalents.

5.2 REMODELAGE

Les groupements que nous venons de proposer sont parfois étonnantes, parfois impossibles aux yeux d'un spécialiste du *Corpus*. Il s'impose en effet que nous les recoupions par d'autres données, en précisant aussi nos positions.

Rien n'est à changer, nous semble-t-il, dans le groupe des "traités rapprochés," puisque l'opinion générale admet que ses 4 sous-groupes doivent sans doute être rattachés deux à deux. Nous sommes un peu surpris de retrouver côté à côté les *Affections internes* (no 42) et la *Nature de la femme* (no 43), car ils avaient été classés par nous dans des catégories différentes (voir point 2.3).

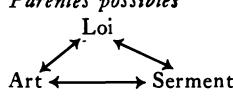
Nous pensons à l'inverse que les 4 ouvrages en "exclusions mutuelles" ont dû être rédigés par 4 auteurs différents, car ils ne sont distingués entre eux malgré le fait que la même catégorie comprenait *Epidémies II*, *Maladies II* et les *Humeurs* (voir point 2.3). Pour simplifier notre classement nous placerons sous le même toit que ces livres ceux que nous avons qualifiés d'isolés: *Prorrhétiques I*, *Foetus de 8 mois* et *Régime salutaire*, car une certaine logique nous dit que, s'ils sont isolés ils s'excluent aussi mutuellement. On notera encore que le *Foetus de 8 mois* et le *Régime salutaire* font partie de la même catégorie de livres (voir point 2.3).

Laissant de côté les ouvrages que nous avons qualifiés de "peu caractérisés," nous devons examiner ceux qui sont rapprochés par des "parentés possibles." Il paraît assez normal que la longueur des mots rapproche des collections de pensées telles que les *Aphorismes* et les *Prorrhétiques*, et en même temps le bon sens interdit de conclure que les deux soient d'un même auteur. On pourrait garder une hypothèse à la vue de ce voisinage: deux compilations de réflexions dues à beaucoup d'auteurs se traduisent en des mots qui se ressemblent par leurs longueurs. D'autre part nous prétendons que la parenté des *Plaies* et des *Hémorroïdes*, de même que celle du *Pronostic* avec les livres I et III des *Epidémies*, n'est que possible parce que les traits qu'ils partagent ne sont pas suffisamment nombreux pour qu'on soit en mesure d'être catégorique à leur sujet.

Nous sommes mal à l'aise devant trois autres ouvrages, car nous avions rangé (voir point 2.3) le *Mochlique* parmi les traités techniques, l'*Officine du médecin* dans les "notes et études" et les *Lieux dans l'homme* dans les "traités médicaux." Ainsi éloignés à cause de leur contenu, ces textes ont malgré cela réussi à mettre en évidence des points communs. Si le *Mochlique* et l'*Officine du médecin* ont tous deux un sommet dans leur courbe à la classe des mots de 2 lettres, et peu de mots très longs, en revanche

TABLEAU 9

Groupements et exclusions parmi 45 traités hippocratique sous l'aspect de la longueur des mots

Traité rapproché	Exclusions mutuelles	Groupe hypothétique
1-Epidémies I Epidémies III	1-Epidémies II 2-Maladies I	Nature de l'enfant Génération
2-Génération Nature de l'enfant	3-Maladies II 4-Humeurs	Affections internes Nature de la femme
3-Fractures Articulations	5-Prorrhétiques I 6-Foetus de 8 mois	
4-Affections internes Nature de la femme	7-Régime salutaire	
5-Mochlique Officine du médecin		<i>Parentés possibles</i> 

l'*Officine* et les *Lieux dans l'homme* ont un second sommet à 6 lettres, une longueur moyenne des mots et un valeur de σ presque identiques. Nous pencherions donc plutôt pour un rapprochement entre ces deux ouvrages. Mais si l'on compare par un *khi* 2 la distribution des mots du tableau 1, en faisant une seule classe des longueurs de 13 lettres et plus, les liens les plus forts apparaissent entre le *Mochlique* et l'*Officine* ($X^2 = 15.85$), puisque entre l'*Officine* et les *Lieux dans l'homme* ($X^2 = 47.38$), puis entre le *Mochlique* et les *Lieux dans l'homme* ($X^2 = 94.81$). Il nous faut donc retenir le groupement *Mochlique—Officine du médecin*, puisque les parentés ne sont plus significatives à 95% à partir de la valeur $X^2 = 21.03$.

Lorsqu'on veut dissocier ou réunir ensuite le *Serment*, la *Loi* et l'*Art*, on a l'impression de tourner en rond, car ils se rencontrent deux à deux sur presque toutes les aspects que nous avons envisagés. On ne saurait expliquer ces ressemblances par la longueur des textes: le *Serment* et la *Loi* sont très courts, mais l'*Art* est 5 fois plus long que les deux autres mis ensemble. S'il est vrai qu'on ne saurait raisonnablement penser que ces trois textes sont d'un même auteur, on doit admettre que leur contenu exerce une influence marquée sur la ressemblance de répartition de la longueur des mots.

Il reste à reviser le "groupe hypothétique" formé de 6 livres, où la sympathie évidente entre les *Plaies de la tête* et le groupe *Génération—Nature de l'enfant* est bien encombrante pour nous, car l'un a la réputation d'appartenir à une école différente des deux autres. La *Nature de l'homme* est réputée coaque elle aussi, mais comme ses caractères sont moins marqués, elle est un peu moins gênante. Le groupement *Affections internes—Nature de la femme*, à son tour est réputé cnidien. Comme nous avions classé ces 6 ouvrages dans 5 groupes différents (voir point 2.3)

quant à leur contenu, il faudrait penser que les œuvres sont à ce point si peu marquées par leur école que même la longueur de leurs mots se ressemble. Nous éliminerons donc les *Plaies de la tête* et la *Nature de l'homme*, puisque l'opinion générale des hippocratisants y voit des traités coaques, sinon hippocratiques. Il nous resterait à expliquer pourquoi ces livres ressemblaient tant aux quatre autres alors qu'ils ne sont pas de la même école ni du même type.

L'ensemble des groupements et des exclusions que nous proposons en définitive sera quelque peu modifié, comme on le voit dans le tableau 9.

5.3 CONCLUSION GÉNÉRALE

Les spécialistes de l'œuvre d'Hippocrate seront en mesure d'apprécier la valeur des rapprochements et des exclusions que nous avons proposés, dont plusieurs correspondent à des faits notoires, d'autres constituant des suggestions nouvelles. Nous n'avons pas la prétention de croire qu'une étude aussi sommaire sur la longueur des mots ait rapporté des résultats inébranlables; il faudrait toutefois être téméraire pour n'en pas tenir compte. A la suite de cette expérience que nous avons menée nous serions heureux si le lecteur avait trouvé, comme nous, que la mesure de la longueur des mots des textes hippocratiques est un instrument utile au regroupement de ces œuvres, car au-delà des impressions personnelles elle fournit une base objective à l'interprétation.

UNIVERSITÉ LAVAL, QUÉBEC

ANNEXE I Liste des livres étudiés

Litré	Numéro d'ordre	Traité du corpus
Tome I	1	De l'Ancienne Médecine
Tome II	2	Des Airs, des eaux et des lieux
	3	Pronostic
	4	Régime dans les maladies aiguës
	5	Appendice aux maladies aiguës
	6	Epidémies I
Tome III	7	Epidémies III
	8	Des plaies de tête
	9	De l'officine du médecin
	10	Des fractures
Tome IV	11	Des articulations
	12	Du Mochlique
	13	Aphorismes
	14	Serment
	15	La loi

ANNEXE I (Suite)

Littré	Numéro d'ordre	Traité du corpus
Tome V	16	Epidémies II
	17	Epidémies IV
	18	Epidémies V
	19	Epidémies VI
	20	Epidémies VII
	21	Des humeurs
	22	Prorrhétiques I
	24	De l'art
	25	De la nature de l'homme
	26	Du régime salutaire
Tome VI	27	Des vents
	28	De l'usage des liquides
	29	Des maladies I
	31	Des lieux dans l'homme
	32	De la maladie sacrée
	33	Des plaies
	34	Des hémorroīdes
	35	Des fistules
	36	Du régime I
	37	Du régime II
Tome VII	38	Du régime III
	39	Du régime IV ou des songes
	40	Des maladies II
	41	Des maladies III
	42	Des affections internes
	43	De la nature de la femme
	44	Du foetus de sept mois
	45	Du foetus de huit mois
	46	De la génération
	47	De la nature de l'enfant