

Approche lexicométrique des discours de Lysias

par

J.-Ph. MASSONIE

M.-M. MACTOUX

Université de Franche-Comté, Besançon - FRANCE

Cette approche lexicométrique du *corpus Lysiacum* est la première étape d'une recherche (1) sur les discours des orateurs attiques pour tenter d'établir une typologie. Nous n'entendons proposer ici que l'état provisoire de ce travail qui, en dehors de son intérêt méthodologique, a une finalité heuristique dans le domaine de l'histoire qui est le nôtre. Nous ne faisons pas de la lexicologie mais voulons introduire, dans une analyse comparative du langage comme fait social, la signification lexicale d'un texte par rapport à d'autres textes, non seulement pour un corpus donné mais aussi pour d'autres corpus offrant, avec le premier, une homogénéité minimale quant à la langue, au lieu, au moment, aux conditions de production et de communication. C'est le cas des plaidoyers prononcés à Athènes entre le dernier quart du Ve siècle et celui du IVe siècle avant notre ère.

Le corpus que nous avons commencé à traiter est constitué par les 35 discours diversement rattachés au nom de Lysias et concernant des événements (2) qui se sont déroulés entre 410 et 380. Totallement enregistré (3), il est décrit par les 368 lemmes les plus fréquents, depuis l'article, avec 6446 occurrences, jusqu'au substantif *χωρίον* emplacement, pays, qui, avec 20 occurrences, est employé dans 8 discours. Les lemmes retenus, tout en couvrant de façon variable chacun des discours les représentent, cependant, d'une manière satisfaisante (4).

Nous avons utilisé l'Analyse Factorielle des Correspondances.

Plutôt que de faire une présentation technique, nous voudrions rappeler le sens théorique de notre démarche. Nous reformulerons ici quelques principes de base qui semblent très évidents, mais nous nous sommes souvent aperçus qu'ils étaient fort négligés.

LE PREMIER POINT consiste à définir l'unité statistique de base.

Ici, l'unité statistique de base est le LEMME. Ce pourrait être le mot (la forme); mais, qu'est ce qu'un mot ? "Secrétariat d'Etat aux Universités" constitue-t-il un mot ou plusieurs mots ? La définition de l'unité statistique sur laquelle nous travaillons doit être donnée de manière claire, en fonction du problème que l'on se pose.

LE SECOND POINT concerne la définition des unités de découpage du corpus.

Là encore, le choix est fonction du problème que l'on se pose. Ce peut être des textes tout entiers, des chapitres, des paragraphes, des phrases ou même des découpages plus fins : des vers, par exemple. Il faut avoir conscience que des problèmes d'échelle peuvent se poser et que tel caractère, qui serait significatif à un niveau, ne le serait pas à un autre.

LE TROISIEME POINT est sujet de bien des querelles : il s'agit de savoir si l'on doit utiliser les occurrences directement ou si l'on doit les corriger en tenant compte de la longueur des unités de découpage ?

Ici, nous avons suivi les deux directions en pensant que les méthodes d'Analyse des Données sont avant tout des méthodes expérimentales et que, lorsqu'un choix ne s'impose pas, il faut utiliser les

deux méthodes et les comparer.

LE QUATRIEME POINT concerne aussi un sujet qui donne lieu à bien des querelles : faut-il prendre tous les mots ou tous les lemmes, ou faut-il se restreindre à ceux pour lesquels les occurrences sont assez fortes ?

Nous pensons, là aussi, que l'intérêt d'utiliser un ordinateur est précisément d'expérimenter et donc de faire des essais en faisant varier la barre pour les occurrences.

LE CINQUIEME POINT concerne l'utilisation de la méthode statistique (ici, l'Analyse Factorielle des Correspondances).

Face aux graphiques d'Analyse Factorielle des Correspondances, il faut respecter deux étapes dans le travail :

La première étape -que nous appellerons "Analyse"- consiste à dire tout ce que l'on peut sur le graphique en ne tenant compte que des propriétés géométriques liées à la méthode, sans introduction du sens des points (individus ou caractères), donc sans apport extérieur d'information.

La deuxième étape consistera à rétablir le "sens" de chacun des points (tel individu, c'est en fait tel texte; tel caractère est en fait tel mot, qui peut avoir tel sens dans le contexte). Cette phase -que nous appellerons "interprétation"- doit s'effectuer en respectant ce qui a été dit dans l'analyse, mais aussi en revenant sans cesse au texte. Il ne doit jamais y avoir de contradiction : toute exception à cette règle doit être justifiée. Faute de respecter ces règles, de nombreuses Analyses des Données ne sont que de vagues discours auxquels l'ordinateur donne une caution, mais sans pour autant en faire un travail véritablement sérieux !

Nous voudrions terminer par quelques remarques sur les rapports qui existent entre l'Informatique et la Statistique.

Dans le cas d'analyse de textes, l'Informatique va permettre de créer des fichiers, de les gérer, de les corriger, puis de poser des questions aux fichiers. Ou ces fichiers sont très petits et, dans ce cas là, point n'est besoin d'ordinateur; ou ces fichiers sont très gros et l'on risque, soit d'être dépassé par le nombre de questions possibles (la combinatoire est l'ennemie), soit de poser des questions avec des *a priori*, et, donc, de se conforter dans ses théories : ce qui est rassurant mais peu scientifique.

La statistique, au contraire, doit être prise comme une méthode permettant de détecter des questions intéressantes en sachant où les poser dans le fichier. Il y a deux approches possibles et qui doivent être explorées successivement :

Soit l'approche au coup par coup : on va détecter des mots ayant des occurrences très fortes, ou répétées en rafale, ou dispersées, etc... et l'on va revenir au texte pour contrôler, ou essayer d'établir des corrélations avec des faits similaires pour d'autres mots, mais l'on sera très vite débordés par ces

corrélations (si on a 10 mots, cela fait 45 coefficients de corrélation à contrôler).

Soit par une méthode d'Analyse des Données, essayer d'avoir une vision globale de l'ensemble de ces corrélations, étant entendu que, avant et après, on étudiera les mots au coup par coup. Le reproche de la lourdeur et du coût d'une telle pratique ne tient plus avec l'apparition des micro-ordinateurs (5).

Ainsi, nous ne sous-estimons pas le caractère réducteur de la manipulation subie par le vocabulaire puisqu'aux mots ont été substitués des lemmes utilisés comme unique descripteur. Mais nous formulons l'hypothèse que ces lemmes sont discriminants à un premier niveau d'analyse et qu'ils sont une voie d'accès au langage comme objet historique, c'est-à-dire comme objet construit. Lemmes, il faut le préciser, qui comprennent -dans cette première analyse- les mots lexicaux et les mots-outils, articles, pronoms personnels, pronoms possessifs, prépositions, etc . . . qui introduisent une syntaxe minimale. Le lemme n'est pas une forme pure, mais le lieu de convergence d'une série d'informations. L'analyse des correspondances (6) qui permet de balayer la surface discursive de 57.321 mots correspondant à 2.769 lemmes et met en rapport inventaires fréquentiels et textes fournit des signes pertinents (7), même si le vocabulaire n'est qu'un aspect, que certains pourront juger mineur, du langage.

La configuration obtenue pour le plan 1/2, le seul sur lequel nous avons, pour l'instant, travaillé, est très particulière. Il apparaît, au vu des graphiques (8), que l'information majeure ne se confond pas avec des axes factoriels qui seraient nettement différenciés. Les axes 1 et 2 ne sont que les modalités d'une structure principale qui s'inscrit sur le graphique d'une manière oblique. Peut-on aller plus loin et proposer une interprétation ? Oui, à condition de s'en tenir à une lecture interprétative qui ne se veut pas, pour l'instant, explicative (9). Ce n'est pas la valeur sémantique du vocabulaire qui organise la structure mais le système énonciatif, ou plus exactement, certains de ses éléments jouant ici un rôle prédominant. On pourrait, en effet, voir sur l'axe 1 du graphique (10) une opposition grossière entre, à gauche, le vocabulaire de la famille, du "je", du "toi", du "nous" . . . en un mot des relations interpersonnelles, et, à droite, le vocabulaire de la cité, de la *polis*, relations à l'intérieur de la cité au plan politique, juridique, religieux et relations des cités entre elles avec leurs deux composantes, les alliés et les ennemis. Au vocabulaire des rapports naturels s'opposerait le vocabulaire des rapports construits.

Mais on peut affiner cette interprétation grâce à la recherche des termes suremployés en commun par un groupe de discours proches. Le repérage de ces termes a été réalisé par un programme statistique qui fournit les termes sur-employés et sous-employés par rapport à la moyenne d'emploi dans l'ensemble du corpus normé. En rapprochant ce vocabulaire du vocabulaire propre à un groupe de discours voisins, c'est-à-dire proche sur les axes, il se dégage un trait commun et par là même signifiant. Ainsi, pour la configuration de l'axe 1, le vocabulaire suremployé en commun par les 3 discours les plus éloignés à droite sur cet axe, les discours 2, 33, 34 (11), se compose de 3 termes : le substantif *κίνδυνος*, danger, et deux particules bien connues en grec et très souvent en relation, *μέν*, *δέ*. C'est un vocabulaire essentiellement référentiel, qui renvoie non pas aux instances du discours mais au référent extra-linguistique. *Μέν* et *δέ* qui servent dans la plupart des cas à créer un balancement antithétique plus ou moins

marqué montrent ce dont on parle. Ces termes ont la même fonction discursive que les appellatifs "ancêtres", "Lacédémoniens", le substantif "salut", ou les verbes "combattre", "en venir à", "se ressouvenir", qui sont les plus éloignés à droite sur l'axe 1 ; donc ceux qui sont les plus spécifiques du groupe de discours considérés, au terme de l'analyse des correspondances, et participent avec eux très fortement à la signification de l'axe.

Je prendrai un autre exemple avec l'ensemble des trois discours qui sont à l'opposé sur l'axe 1, c'est-à-dire à l'extrémité gauche de cet axe : les discours 17, 23, 8.

Le vocabulaire proche est composé de verbes performatifs même si leur emploi ne correspond pas à l'usage performatif au sens strict (12). Ces verbes tels "juger", "témoigner", "appeler" (13) qui tirent leur caractère performatif du code juridique auquel ils appartiennent, sont très différents du verbe "combattre", à l'opposé sur l'axe 1. Ce ne sont pas des verbes qui décrivent une action qui se déroule ailleurs, mais des verbes dont l'énonciation est un moyen d'accomplir cette action. Cette valeur énonciative est confirmée par celle des trois termes suremployés en commun par les trois discours précédents : sembler, avoir l'impression, d'abord; deux verbes et un adverbe qui modalisent, introduisant la marque du locuteur dans l'énoncé centré, de ce fait, sur l'émetteur et non sur le référent. Autrement dit, et l'essentiel du vocabulaire situé le long de l'axe 1 le confirme, cet axe correspond à un axe référentiel. Les discours se classent, de droite à gauche, en fonction d'un vocabulaire de moins en moins orienté vers le référent et de plus en plus sur l'émetteur. A droite, on a des termes qui montrent le monde, le racontent, caractérisant des discours appartenant au registre narratif. A gauche des mots qui, renvoyant aux instances du discours, commentent le monde plus qu'ils ne le racontent.

En simplifiant, on peut dire, qu'aux discours narratifs s'opposent les discours non-narratifs (14) qu'on pourrait appeler énonciatifs. Organisation, qui, à la fois, recoupe un classement par genre oratoire et le dépasse très largement. A l'extrême droite, par rapport à l'axe 1, on trouve les deux discours épidiectiques : *Discours Olympique* (33), *Oraison Funèbre* (2) et le seul discours attribué à Lysias et donné par les Grecs (15) comme délibératif : *Contre une proposition tendant à détruire à Athènes, le gouvernement transmis par les ancêtres* (34), le délibératif et l'épidictique étant reconnus comme différents du judiciaire (16). Mais ils ne s'opposent pas aux discours judiciaires selon cette répartition traditionnelle. Parmi ces derniers, un certain nombre sont beaucoup plus narratifs que d'autres, les causes privées traitées par les discours 17, 23, 10, 11, auxquels peut se rattacher le 8, qui n'est pas à proprement parler un plaidoyer, représentant les discours les plus énonciatifs. L'analyse des correspondances fait éclater la vision commune (17).

L'axe 2, quant à lui, est organisé par les pronoms personnels qui n'accompagnent pas les verbes grecs sous la forme sujet mais seulement complément, sauf pour marquer l'insistance, et par les adjectifs-pronoms possessifs des trois personnes. De bas en haut, le long de cet axe, on rencontre le tu désignant l'adversaire, le je, moi, mon, renvoyant au locuteur-plaideur, situés dans la partie inférieure; et dans la partie supérieure, encore proches du centre, des pronoms mixtes, le nous, le nôtre équivalent

à moi + eux, famille du locuteur ou Athéniens, le pronom eux, non réfléchi et réfléchi, qui renvoie essentiellement à certaines catégories d'Athéniens, et enfin, à l'extrémité, le vous, le vôtre, qui désignent d'une manière quasi exclusive les Athéniens.

A un échange verbal dominé par le rapport locuteur-plaideur/adversaire, qu'il s'agisse de discours prononcés par les demandeurs ou les défendeurs, répond un autre type de communication discursive privilégiant le deuxième interlocuteur d'une relation triangulaire constituée par les deux adversaires et les juges, membres des tribunaux, et les Athéniens en général entre lesquels s'établit une coïncidence : en effet, à l'extrémité supérieure de l'axe 2, se côtoient non seulement le pronom vous, mais aussi le substantif juges, et des noms tels Athéniens, peuple, sous ses deux formes, *δῆμος* et *πλῆθος*, très souvent utilisé avec l'adjectif possessif votre, "votre peuple", accompagnés de verbes tels condamner à mort ou son antonyme, acquitter, qui ont pour sujet le vous non exprimé, c'est-à-dire les Athéniens-juges.

Quant aux termes suremployés, les discours 15, 27, 28, qui sont les plus éloignés sur l'axe 2 dans sa partie supérieure, n'en ont pas en commun. Ils figurent seulement dans ces discours reclassés par groupe de deux : 15 et 27; 15 et 28. Mais leur signification va dans le même sens. Ils dénotent la 3e personne, soit qu'ils la nomment comme *ἄνθρωπος*, l'homme par rapport à la femme mais aussi celui qui est chargé d'une fonction publique, stratège, épheure ou juges; *ἄνθρωπος* est très souvent lié dans les textes à *δικαστής*, juges; soit qu'ils la désignent comme la particule *ὧς* avec laquelle on s'adresse à quelqu'un; soit qu'ils impliquent la nécessaire création d'un consensus entre les interlocuteurs, au plan des actions ou au plan des sentiments, avec des verbes ou expressions verbales tels *χρή*, il faut, ou *δεωός* exprimant le caractère inadmissible d'une situation qu'il faut changer.

Ainsi cet axe 2 me paraît définir le mode de communication : à un mode communicatif caractérisé par un échange entre le locuteur-plaideur et son adversaire, se substitue un autre mode dominé par la relation discursive qui privilégie la communauté civique. Les destinataires n'ont pas, dans tous ces discours, le même rôle discursif.

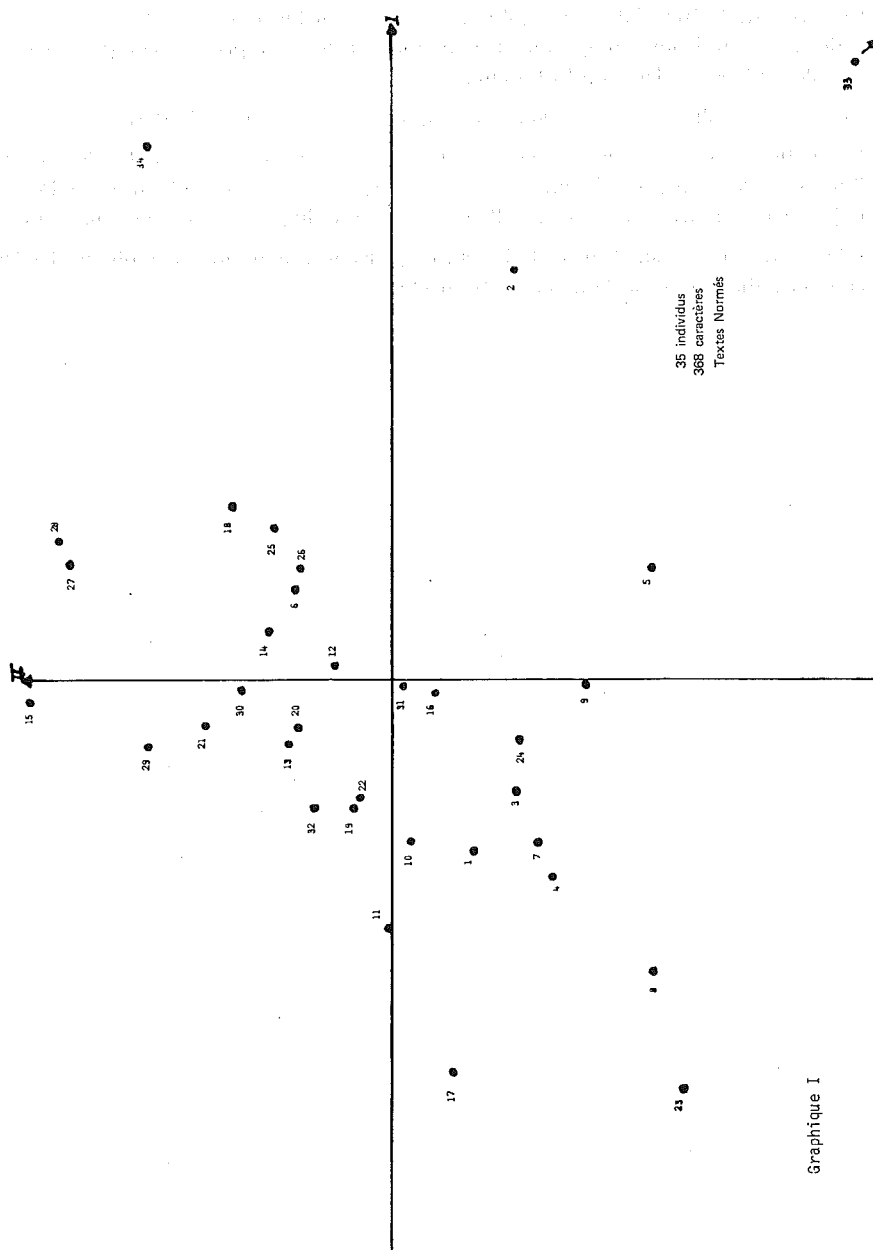
En conclusion nous dirons que les axes factoriels 1 et 2 permettent de saisir deux modalités voisines mais différentes de la structure énonciative des discours du *corpus Lysiacum*. D'autres analyses sont nécessaires; celle, par exemple qui supprimerait les lemmes à référence personnelle pour voir si une moindre dispersion entraînerait corrélativement un meilleur contraste entre les discours. L'analyse des correspondances est, avant tout, une méthode expérimentale qui, en multipliant les prises de vue sur un même objet, suscite de nouvelles questions qui ne pourront être résolues que par un retour incessant aux textes.

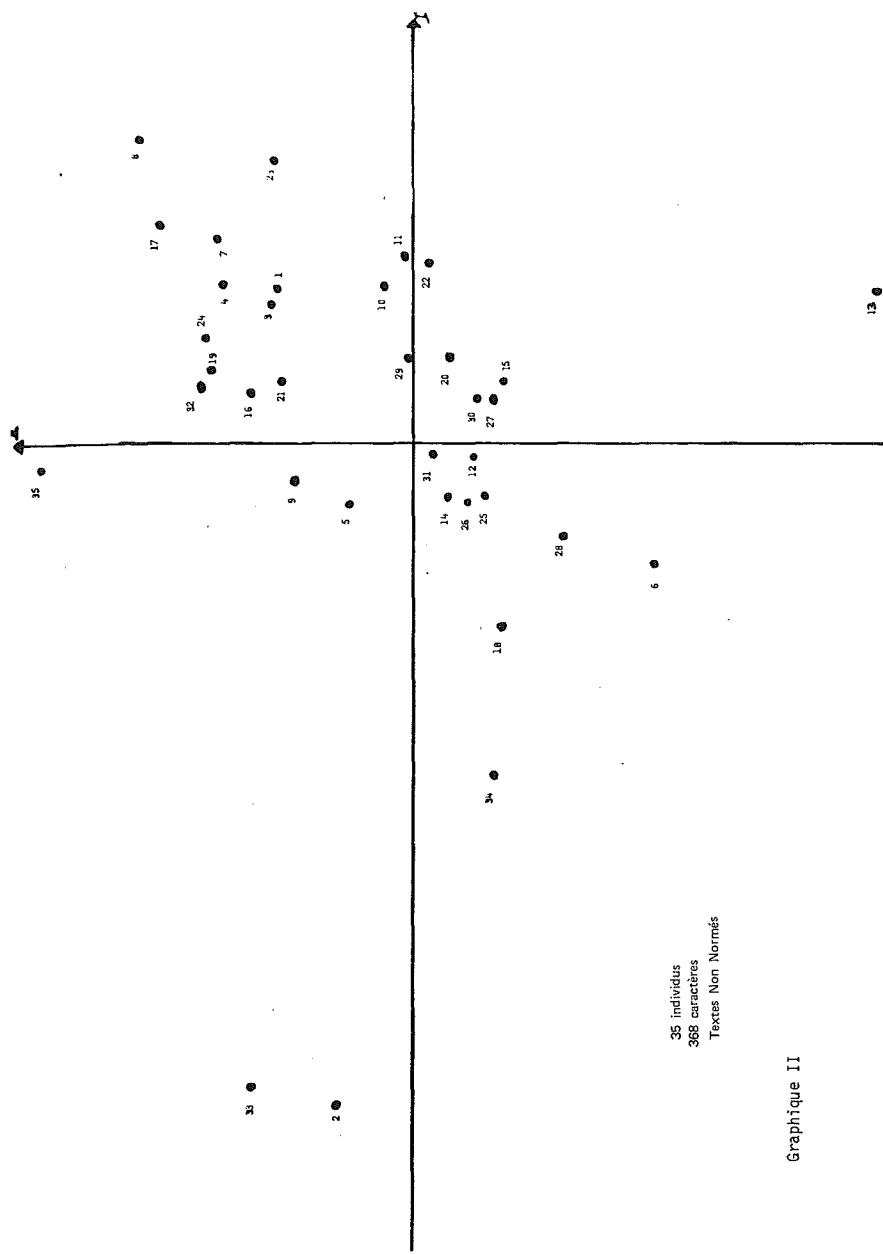
NOTES

- (1) Elle est menée dans le cadre de l'ERA 520, Esclavage et formes de dépendance dans l'Antiquité, et du GIS 127777 du CNRS, *Techniques nouvelles en Sciences de l'Homme*.
- (2) Personnels dans le cas des procès, ou collectifs pour les discours épidiectiques.
- (3) Par le L.A.S.L.A. et lemmatisé à l'aide de son programme. L'enregistrement d'Antiphon est en cours de vérification; celui d'Andocide en voie de réalisation. Le prochain orateur attique enregistré sera Isocrate.
- (4) De 73 à 94 %.
- (5) Avec un système de 25.000 FF, on peut -sans grande difficulté- décortiquer un corpus équivalent, en volume, à un gros roman.
- (6) A partir d'un tableau de contingence, avec, en colonne, les 35 discours et, en ligne, le nombre d'occurrences des lemmes pour chaque discours.
- (7) Cf. la série d'études menées ces dernières années in J.P. BENZECRI et Collaborateurs, *Pratique de l'analyse des données*, 3, *Linguistique et lexicologie*, Paris, 1981.
- (8) La remarque vaut pour les graphiques I et II donnés en annexe et correspondant aux textes normés et non-normés. Nous n'aborderons pas ici le problème de la norme. Dans le cas très précis des discours de Lysias, le fait de normer ne modifie pas de manière fondamentale, la structure générale. Les discours les plus éloignés du centre demeurent les mêmes.
- (9) La multiplicité des facteurs explicatifs invite à la prudence. La différence d'auteurs, même pour un corpus dont la quasi-totalité (sauf le XII qui traite d'une affaire personnelle) a été récemment refusée à Lysias comme unique rédacteur (Cf. K.J. DOVER, *Lysias and the corpus lysiacum*, Berkeley and Los Angeles, 1968) n'est pas nécessairement la plus décisive. Sur nos graphiques, le discours XII est très près du centre. Il apparaît aussi, du point de vue lexical, comme relativement proche de la moyenne des 35 discours; autrement dit, la totalité des discours ont un profil lexical qui ressemble pour une part à celui du XII.
- (10) Nous nous appuyerons désormais sur le graphique I comportant à la fois les points-discours et les points-lemmes. Il ne nous a pas été possible de le joindre dans son intégralité.
- (11) Les numéros sont ceux de la collection des Universités de France.
- (12) A la première personne du singulier du présent. J.L. AUSTIN, *Quand dire c'est faire*, Paris 1970, à qui cette notion est empruntée, envisage lui-même son extension.
- (13) Sa nature performative est renforcée par son emploi, dans plus de la moitié des cas, à l'impératif, pour appeler les témoins. Il y a coïncidence, pour le locuteur, entre la forme verbale et l'action signifiée par elle.
- (14) Cf. H. WEINRICH, Les temps et les personnes, *Poétique*, 1979, 39, p. 338-352, qui cherche à fonder, sur la syntaxe, l'opposition narratif et non-narratif. Il conviendrait certes, d'examiner

les autres traits énonciatifs. Ce qu'on peut dire, pour l'instant, c'est qu'il y a une différenciation lexicale caractéristique, qui, n'importe comment, devra être prise en compte par une analyse globale du langage dans son historicité.

- (15) Cf. Denys d'Halicarnasse, *Les orateurs attiques, Lysias*, 31, I, Paris 1978.
- (16) Distinction proposée pour la première fois par Aristote, *Rhétorique I*, 1358b. 1-8 et sur laquelle il y aurait des remarques à faire. Dans son analyse le délibératif a bien des points communs avec le judiciaire. Dans les deux genres l'auditeur est un "juge" que l'on conseille ou déconseille.
- (17) Elle repose, certes, sur d'autres traits qui ne peuvent être atteints à ce niveau d'analyse. Mais cette classification lexicale ne peut être négligée.





35 individus
 368 caractères
 Textes Non Normés

Graphique II