

ANALYSE THÉMATIQUE AUTOMATIQUE

par LOUIS DELATTE

J'ai choisi de vous parler d'une étude qui est en cours de réalisation dans mon Laboratoire (LASLA) et qui, d'une certaine façon, je crois, rejoint les préoccupations de nos hôtes. En effet, non seulement elle fait appel aux techniques que nous avons mises au point pour la lexicalisation des oeuvres antiques, latines et grecques, mais encore elle tente, par le moyen de programmes complexes, d'analyser automatiquement le contenu sémantique des textes que nous avons enregistrés en ordinateur.

Avant d'aborder cette étude, et pour rendre la suite de l'exposé plus compréhensible, il me faut dire un mot des travaux que nous avons entrepris depuis 13 ans et décrire rapidement les techniques auxquelles je viens de faire allusion.

Ce Laboratoire s'est donné pour tâche essentielle d'analyser les textes anciens sous leur aspect le plus formel. Grâce à de telles analyses, nous espérons résoudre des problèmes que la philologie classique traditionnelle a été impuissante à résoudre: il s'agit notamment des problèmes d'authenticité et de chronologie relative des oeuvres d'un auteur. Nous croyons aussi que les éléments recueillis par ces analyses nous permettront de comprendre les mécanismes profonds de la création littéraire à tous les niveaux, même les plus inconscients.

Pour atteindre ces objectifs, nous nous fondons sur l'étude statistique des mots d'un texte aux points de vue phonétique,

morphologique, lexicologique, syntaxique et stylistique, ce qui suppose l'établissement de fichiers extrêmement complets.

Leur réalisation et leur exploitation exigent naturellement le recours aux machines mécanographiques et à l'ordinateur.

Nous avons donc été amenés non seulement à imaginer une technique pratique pour codifier nos informations sur cartes mécanographiques, mais encore à prévoir tous les éléments d'analyse formelle qui nous permettraient de répondre aux diverses questions auxquelles je viens de faire allusion.

Quelles informations transcrivons nous sur cartes et comment?

La base de tous nos travaux, c'est l'oeuvre littéraire antique qui se présente à nous sous l'aspect formel d'une suite définie de mots. C'est donc chacun de ces mots qui est traduit dans ce « langage-machine » que constituent les perforations, en prenant comme principe de n'écrire qu'un seul mot par carte. Mais comme ces cartes-mots seront mélangées à la suite de divers tris ou par accident de manipulation, comme d'autre part, le nombre de ces cartes va atteindre très vite de grandes proportions, il est essentiel que chaque mot soit accompagné de références extrêmement précises qui permettront de le retrouver et de le reclasser facilement. Chaque mot est donc situé, d'abord par son numéro d'ordre dans l'oeuvre, par le numéro du chapitre, le numéro du paragraphe ou du vers, puis le numéro d'ordre dans le paragraphe ou dans le vers et enfin le numéro d'ordre dans la phrase.

Ces références sont perforées automatiquement dans chaque carte par une calculatrice. Elle se base sur un code spécial perforé en colonne 65, qui, par convention, signifie fin de phrase, fin de chapitre, fin de paragraphe. Ce système de références, outre qu'il nous est utile pour l'étude statistique des longueurs de phrases, pour l'étude des clausules métriques, pour une étude stylistique des mots qui terminent une phrase ou un vers,

nous met à l'abri de tout accident. En effet, qu'une carte soit déplacée ou égarée dans un fichier, un contrôle de séquence portant sur le numéro d'ordre dans l'oeuvre détectera l'erreur et signalera immédiatement que l'ordre est perturbé ou qu'un élément de la série est manquant.

A ce stade, nous n'avons encore sur la carte que la référence et la forme du mot tel qu'il apparaît dans l'oeuvre. Or, certaines formes d'un mot sont si différentes les unes des autres que la mise en ordre alphabétique serait impossible, si on ne les regroupait pas toutes sous un même lemme, c'est-à-dire sous la forme du mot tel qu'il figure au dictionnaire. Comment, par exemple, regrouper les différentes formes de *esse* (*sum, eram, fui*), si l'on n'a pas le lemme *sum* comme référence générale?

A défaut d'une telle organisation, une liste de mots-formes est pratiquement inutilisable. Il faut donc que nous transcrivions sur la carte le lemme de la forme: pour un verbe, par exemple, ce sera la première personne du singulier de l'indicatif présent actif; pour un substantif ou un adjectif, la forme du nominatif singulier, etc.

La carte porte maintenant la forme du mot telle qu'elle apparaît dans le texte, le lemme et la référence.

Une nouvelle difficulté se présente ici: c'est le problème des homographes. Si l'on se réfère à un lemme sans autre indication plus précise, beaucoup de mots et de formes s'écrivant de la même façon, mais n'ayant pas le même sens ou la même construction, se confondent dans les listes. Après divers tâtonnements, nous avons décidé d'affecter les lemmes homographes d'un indice numérique: ainsi *tempus 1*, le temps, et *tempus 2*, la tempe.

Pour nous qui, dès avant la création du Laboratoire, voulions faire une analyse très poussée des textes, le grand problème a été de donner à propos de chaque mot le plus grand nombre de renseignements possible et de ne rien oublier, puisque, une fois le travail entamé, il serait bien difficile de revenir en

arrière. Après d'inévitables hésitations, nous avons résolu de faire une analyse morphologique et syntaxique complète de chaque forme.

Toutes ces analyses sont transcrites dans un code alphanumérique qui permet de condenser l'information. Vous avez d'ailleurs sous les yeux, nos codes d'analyse morphologique¹. Ainsi, pour ne prendre qu'un exemple très simple,

le substantif est codifié par 1 en colonne 54,

la 1ère déclinaison est codifiée par 1 en colonne 55,

le verbe est codifié par 5 en colonne 54,

la 1ère conjugaison est codifiée par 1 en colonne 56.

Voici d'ailleurs la reproduction d'une carte-mot complète.

Tout le travail d'analyse qui, pendant 4 ans, a été fait artisanalement était fort lent: un philologue ne pouvait guère analyser plus de 250 mots par jour.

Nous avons alors décidé, riches de notre expérience, d'automatiser l'analyse morphologique. Ce programme, qui nous a pris 6 mois de travail, analyse le latin à raison de 10.000 mots par nuit sur notre propre ordinateur. Sans doute, tout le travail de la machine doit être revu par le philologue, en raison du fait que l'ordinateur sort toutes les analyses possibles pour chaque mot. Mais il simplifie étonnamment la tâche du philologue qui peut très aisément analyser 2000 mots par jour.

Ces quelques explications vous montrent sur quelles bases et avec quels outils nous poursuivons nos recherches.

J'en viens maintenant à l'objet propre de cette communication.

Une des tendances les plus marquantes de la critique actuelle est constituée par la critique thématique: l'étude des différents thèmes qui sous-tendent l'oeuvre littéraire et l'analyse

¹ Cf. annexe 1.

de la manière dont ils s'articulent les uns aux autres doivent permettre d'atteindre la structure affective particulière qui donne à l'oeuvre sa cohérence et son unité.

Pour cette nouvelle critique, le thème revêt donc un sens particulier. Mais il ne s'agit plus, comme pour la critique traditionnelle, d'un sujet traité par l'auteur volontairement et en pleine conscience. Au contraire, il s'agit, comme le dit R. Barthes, d'un « réseau organisé d'obsessions » relevant sinon de l'inconscient, du moins des zones extraréflexives de la conscience. Structure fondamentale de l'imagination créatrice, le thème ne s'exprime pas directement dans l'oeuvre, mais se trahit par des constantes formelles, des figures obsédantes, des réseaux associatifs stables souvent masqués par les structures voulues du texte.

C'est là précisément que se situe, pour le critique, la principale difficulté: comment détecter les affleurements de ces thèmes? Faute d'une méthode rigoureuse d'investigation, les recherches thématiques comportent une part trop grande de subjectivité, chaque critique ayant tendance à privilégier certains thèmes en fonction de son équation personnelle.

L'étude que nous avons entreprise vise précisément à faire apparaître par des moyens automatiques et, par conséquent objectifs, les associations de thèmes à l'intérieur d'une oeuvre. Le réseau que l'ordinateur permet de construire conduit ainsi à l'élaboration d'un schéma structural de l'ensemble.

Toute la recherche repose sur une exploration automatique de contextes dont la longueur peut être variable et est déterminée, en début de travail, par le philologue ou le critique.

D'une manière simplifiée, on peut considérer qu'elle répond en fait à cette question: lorsque, par exemple, dans un texte, je rencontre l'un ou l'autre des mots qui font partie du champ notionnel « famille », quels sont les mots d'autres champs notionnels qui se trouvent dans le voisinage immédiat des premiers et avec quelle fréquence se rencontrent-ils?

Les associations auxquelles je viens de faire allusion sont

donc — il importe d'y insister — des associations purement formelles, sans qu'elles présentent nécessairement entre elles de liens logiques.

Mais nous croyons que c'est précisément à ce niveau que nous pourrions mettre au jour ces thèmes inconscients, ces ressorts cachés qui confèrent à une oeuvre sa véritable signification.

Dans la suite de cet exposé, on définira *le champ notionnel* comme l'ensemble des lemmes qui désignent un même concept, une même notion ou qui présentent avec lui des rapports étroits. Ainsi, les mots *dieu, divinité, déesse, divin*, font partie du même champ notionnel caractérisant le divin. *Un champ thématique* est l'ensemble constitué par un ou des membres d'un champ notionnel associé à un ou des membres d'autres champs notionnels dans la mesure où ces derniers se rencontrent de manière significative dans le voisinage immédiat des premiers.

Le processus d'automatisation décrit ci-après se divise en en trois phases distinctes:

I. Constitution et codification automatique des champs notionnels.

II. Recherche des contextes pour chaque lemme appartenant au champ notionnel étudié; établissement d'un index particulier pour ce champ, puis d'un index général se rapportant au champ thématique constitué par l'ensemble des champs notionnels.

III. Recherche des connexions entre champs thématiques.

I — *Constitution d'un fichier de champs notionnels sur cartes perforées.* — Ce fichier contient tous les lemmes de l'oeuvre étudiée accompagnés de leur traduction et d'un code numérique à valeur sémantique. Ainsi, une carte sera réservée à *pater* accompagné de sa traduction *père* et affecté de 2 codes numériques: l'un indique sa fréquence, l'autre représente le champ *famille*: 055. (Cf. annexe 2).

Pour constituer ce fichier, nous avons fait appel à deux

listes: la première est la liste de fréquence telle qu'elle a été établie par l'ordinateur à la suite de la constitution de l'index de l'oeuvre. (Cf. annexe 3).

Chaque lemme reçoit la traduction de base que lui donne le dictionnaire de Gaffiot pour le latin et le dictionnaire de Bailly pour le grec. On pourrait objecter que cette traduction ne recouvre pas les sens divers que peut prendre un mot, en particulier les sens dérivés par usure de la langue ou par effet de style et que, par conséquent, pour certains mots, l'attache à tel champ notionnel est assez artificielle. Mais cette objection est surtout valable pour les mots-thèmes, tels que *pouvoir* ou *savoir*, *faire* ou *dire*, qui ont une grande extension de sens, mais n'ont guère de valeur sémantique. Elle est de peu de poids pour les mots de caractérisation qui sont d'une grande restriction de sens et très précis et qui, par définition, sont les seuls mots réellement intéressants pour une étude thématique. D'ailleurs, afin de limiter la recherche aux seuls mots pleins, on négligera la traduction de deux catégories particulières de lemmes: les noms propres (sauf si l'on désire les faire intervenir dans les regroupements ultérieurs) et les mots-outils (conjonctions, prépositions, verbe copule, etc.). Ces lemmes sont caractérisés par un code spécial: N pour les noms propres et O pour les mots-outils.

La seconde liste est, en fait, un dictionnaire de notions, par exemple: le divin, l'amour et la haine, la famille, etc. Ce dictionnaire a été réalisé artisanalement en procédant à des regroupements fondés aussi bien sur les familles étymologiques (père, paternel) que sur les champs sémantiques (père, mère, frère, soeur). On trouvera un échantillon de quelques groupes en annexe 4.

Sans doute, est-ce ici le point faible de la recherche parce que la détermination des mots qui font partie d'un champ notionnel est affaire d'appréciation personnelle, que la définition peut varier selon l'angle sous lequel on envisage la recherche et qu'en conséquence le regroupement de divers mots au sein d'un

champ notionnel reste subjectif. Nous croyons que ce n'est qu'à la suite de multiples expériences et de corrections successives que nous pourrions constituer un Corpus de notions efficaces.

Chaque bloc notionnel a fait l'objet d'un enregistrement sur cartes perforées et a reçu un numéro de trois chiffres établi automatiquement par l'ordinateur.

A partir de ces données, c'est-à-dire le fichier fréquentiel des mots de l'oeuvre et le dictionnaire de notions, l'ordinateur constitue automatiquement un fichier de champs notionnels: il compare le fichier des mots de l'oeuvre à son dictionnaire de notions et il attribue à chaque mot le code numérique qu'il trouve dans son dictionnaire. A la fin du travail, les lemmes de l'oeuvre se répartissent en quatre catégories:

1. Les mots qui font partie des différents champs notionnels.
2. Les noms propres, codifiés N.
3. Les mots-outils, codifiés 0.
4. Les mots indéterminés qui ne portent aucun code, soit parce qu'ils n'ont pas reçu de traduction dans le premier fichier, soit parce qu'ils n'ont pas d'équivalent dans la liste des concepts.

J'ajouterai un mot sur la façon dont s'effectue la codification. L'ordinateur lit la traduction de chaque mot, caractère par caractère: il arrête sa consultation dans deux cas.

1) Lorsqu'il rencontre un blanc ou un signe spécial tel que le tiret.

2) Lorsqu'il a lu les dix premiers caractères. De la sorte, il considère comme un sème unique: accompagner et accompagnement et aussi: placer-devant et placer-auprès de.

Pour les oeuvres que nous avons étudiées jusqu'à présent, l'*Electre* d'Euripide, *Médée*, *Agamemnon* et *Phèdre* de Sénèque, nous avons constitué 125 champs notionnels qui permettent de recouvrir environ 80% des mots pleins, soit 65% de l'ensemble des lemmes de chaque oeuvre. Tous les documents-échantillons que vous avez sous les yeux se réfèrent à la *Médée* de Sénèque.

II - *Recherche des contextes et constitution des Index.* — Lorsque l'ordinateur a achevé son travail de codification et a transcrit le résultat sur cartes perforées, on trie ce nouveau fichier de façon à rassembler tous les lemmes appartenant au même champ notionnel: ils sont aisément reconnaissables puisqu'ils portent le même code numérique.

Il est donc loisible, à ce moment, de choisir les champs que l'on désire étudier. Par exemple, si je veux étudier les associations du groupe « famille » représenté par le code 055 et le groupe « pureté » représenté par le code 004, j'aurai à ma disposition deux jeux de cartes dont les unes portent des lemmes tels que père, mère, fille, etc. et le code 055 et les autres, des lemmes tels que pureté, innocence, chasteté et le code 004. Ce sont ces jeux de cartes qui sont enregistrés en mémoire centrale et ce sont eux qui constitueront pour l'ordinateur un signal d'alerte chaque fois que, à la lecture de l'oeuvre, il rencontrera un mot appartenant à ces champs notionnels.

Naturellement, le travail en machine se fait séquentiellement, thème après thème et tout le traitement décrit ici est exécuté autant de fois qu'il y a des thèmes à étudier, le but de la seconde phase du traitement étant de rechercher et d'imprimer les contextes relatifs à chaque mot faisant partie du thème étudié puis de constituer une liste fréquentielle de tous les lemmes qui apparaissent dans ces contextes. La longueur du contexte est libre et précisée à l'ordinateur préalablement à la recherche proprement dite.

Voici comment s'effectue le travail. L'oeuvre à traiter est enregistrée sur disques dans l'ordre du texte et chaque mot-forme est accompagné de son lemme et de sa référence. Ce sont les cartes issues de l'analyse morphologique automatique, corrigées par le philologue et qui ont servi à l'établissement de l'Index de l'oeuvre. L'ordinateur lit le texte en séquence et constitue, dès le départ, deux matrices qui sont composées d'un nombre d'éléments égal au nombre d'occurrences que l'on souhaite pour

les contextes. La première matrice est réservée aux mots-formes: elle contient donc le contexte proprement dit. La seconde reprend uniquement les lemmes correspondants aux formes et est destinée à la préparation de la liste fréquentielle. Dans l'exemple que nous avons choisi, chaque fois que l'ordinateur rencontre un des mots du groupe « famille », il transfère le contenu des deux matrices dans deux zones d'attente sur disque. Il est important de noter que le programme n'impose à la machine qu'une seule lecture du texte, quel que soit le nombre de mots du champ notionnel. En effet, pour chaque mot du texte qu'elle lit, la machine consulte chaque mot de son fichier notionnel. Lorsqu'elle a achevé l'exploration de l'oeuvre entière, elle procède à trois tris portant sur les fichiers constitués sur disque. Le premier trie le fichier des contextes dans l'ordre alphabétique des lemmes qui appartiennent au champ notionnel étudié; ce fichier est constitué à partir de la matrice I. La seconde matrice permet, grâce à deux tris alpha, de réaliser deux listes fréquentielles. La première liste reprend un des mots du champ, tous les mots qui sont associés à ce champ dans le contexte choisi à *condition qu'ils se présentent au moins deux fois* et un nombre indiquant la fréquence de la rencontre. Ce premier document servira à l'étude psychologique approfondie du thème.

La seconde liste reprend seulement en ordre alphabétique les mots associés à un champ et la fréquence de l'association. C'est cette liste qui va servir de donnée d'introduction pour la troisième phase.

III - *Recherche des connexions entre groupes sémantiques.* — Cette troisième phase consiste dans l'exploitation simultanée du fichier thématique constitué lors de la première phase et du fichier établi pour l'ensemble des groupes lors de la deuxième phase. Le but de cette exploitation est d'attribuer à chaque lemme de ce fichier son propre code thématique. A la suite de cette opération, l'ordinateur possède un fichier sur disque

dont les éléments, réduits à l'essentiel, se présentent de la façon suivante: deux codes thématiques et un chiffre indiquant la fréquence de leur association.

L'ordinateur calcule alors, pour chaque champ notionnel les deux champs auxquels il est le plus fréquemment associé. Ce calcul servira à établir des listes et des tableaux qui visualisent les résultats de l'analyse.

Voici d'abord comment se présente la liste des champs notionnels et de leurs connexions. On y retrouve la traduction des lemmes et les connexions que chacun de ces lemmes présentent avec d'autres champs notionnels. Le premier chiffre indique le champ où le mot est le plus fréquent, le second chiffre indique la fréquence immédiatement inférieure. Cf. annexe 5.

Voici maintenant le tableau qui est une illustration plus synthétique, plus parlante de la liste. Ce tableau est réalisé automatiquement par la machine. Cf. annexe 6. Les codes des champs notionnels se trouvent indiqués aussi bien en abscisse qu'en ordonnée, de sorte que à l'intersection des lignes et des colonnes, l'ordinateur imprime les codes des champs qui sont associés aux champs indiqués en tête de ligne et en tête de colonne. La fréquence des associations a été signalée par la convention suivante: le champ qui se trouve à l'intersection est plus fréquemment employé avec le champ qui se trouve en tête de ligne qu'avec le champ qui se trouve en tête de colonne. Ainsi dans le tableau que vous avez sous les yeux, on constate immédiatement qu'à l'intersection des champs: famille et pureté, on trouve 29 champs différents, ce qui implique fatalement des rapports étroits entre les champs famille et pureté.

Si l'on pousse l'analyse un peu plus loin dans le détail, on constatera, par exemple, que le champ 001, représentant le langage militaire (arme, blessure, épée, frapper etc.) est employé fréquemment dans le contexte immédiat de famille et aussi, quoique avec une fréquence moindre, dans le contexte de pureté. Si le premier champ thématique se comprend aisément dans la tragédie de

Médée, la connexion pureté-arme est moins évidente et demande un retour au texte pour être explicitée.

Ce qui surprendra aussi, ce sont les contextes du champ 061: amour-haine. On pourrait croire, en effet, que *Médée* est une amoureuse trahie et elle est cela, certainement, au début de son aventure: l'on attendrait donc un champ thématique important à propos de cette situation. Mais il n'en est rien: on ne trouve aucun contexte commun avec les lemmes du langage militaire qui sont souvent, cependant, comme le montrent les élégiaques latins, l'expression du langage amoureux. Aucun contexte non plus avec les autres champs notionnels, sauf avec celui de la famille où l'amour prend une autre signification: amour filial, amour maternel.

Tout se passe donc comme si Sénèque avait voulu nous montrer une *Médée* qui a dépassé le stade de l'amour-passion et qui en est au moment ultime de son évolution psychologique, une *Médée* dépouillée et qui n'a plus devant elle qu'une réalité fondamentale: celle de la famille. C'est là que se situe le vrai conflit ainsi que le montre l'abondance des champs thématiques 055-004 et 055-058.

Or, il faut dire qu'une telle conception n'apparaît pas évidente à la lecture de la pièce. L'ordinateur nous livre là une interprétation nouvelle et qui demandera à être explicitée par une étude approfondie du texte et aussi par des comparaisons faites sur d'autres pièces, par exemple *Phèdre*. Sur un plan beaucoup plus inconscient et dans l'optique d'une critique sociologique, on constatera que Sénèque fait porter tout le poids de la pièce sur la notion de famille diffuse à travers toute l'oeuvre. Ne serait-il pas légitime de voir dans cette mise en évidence un reflet des préoccupations du philosophe, anxieux de la dislocation de la famille à son époque?

Il me faut cependant reconnaître que ce tableau, s'il permet une vue synthétique des connexions thématiques, ne rend pas parfaitement compte des rapports qui unissent les groupes, en

raison du fait qu'il ne donne pas d'indication précise sur les fréquences exactes des connexions entre thèmes.

Aussi un dernier groupe de programmes imprime-t-il des listes qui permettront une analyse fine des réseaux associatifs. Ces listes se présentent de la manière suivante: l'ordinateur donne d'abord, pour chaque champ notionnel, le nombre de lemmes et le nombre d'occurrences par lesquelles ce champ est représenté dans l'oeuvre entière. Ensuite, il calcule quels sont ceux de ces lemmes qui sont associés au champ notionnel sous étude et il en donne la liste et le nombre d'occurrences. Cf. annexe 7. Grâce à ces listes, disais-je, le critique peut se livrer à une analyse plus fine parce que la notion de fréquence y apparaît avec plus de précision que dans le tableau général et qu'en conséquence, on peut faire intervenir des tests statistiques, tels que le χ^2 , ou une distribution binomiale pour mettre en relief le degré de pertinence de la rencontre de deux champs notionnels.

Telle est, Mesdames et Messieurs, une des recherches que nous avons entreprises à l'heure actuelle à Liège. Je ne me dissimule pas combien elle est encore imparfaite, en raison principalement des problèmes que pose la constitution d'un dictionnaire notionnel.

Mais, indépendamment de cette question, il est évident que la recherche automatique des thèmes, tout en étant un premier essai d'intelligence artificielle, ne constitue qu'une étape préparatoire à une analyse humaine. Cette étape est cependant indispensable car c'est elle qui déterminera l'orientation des recherches en mettant en lumière des rapports sous-jacents, qui, sans elle, seraient passés inaperçus.

54	55	56	57	58	59	65	66
Catégorie grammaticale	Sous-catégorie Degré Voix	Cas Personne Nombre	Mode	Temps	Fonctions	Tradition	Fonction
Substantif 1	1e décl. 1 2e décl. 2 3e décl. 3 4e décl. 4 5e décl. 5 Anomal 6 Décl. gr. 7				Verbe princ. 12 Verbe subord. 11	Leçon contestée 1 Mot absent dans certains ms. et adopté par édit. 2 Conjecture reçue 3 Lacune comblée CRUX 4 Leçon rejetée 6 Mot présent dans certains ms. et rejeté par édit. 7 Mot impossible à analyser en fonction d'une lacune 8	K dernier mot du chap., du paragraphe ou du vers et de la phrase 8 dernier mot du paragraphe ou du vers et de la phrase 2 dernier mot de la phrase 11 dernier mot du chapitre et du paragraphe ou du vers 0 dernier mot du paragraphe ou du vers
Adjectif 2	P C S 1e class. 1 A B J 2e cl. conc. 2 B K -er 3 C L -is 4 D M Imp. 5 E N Anomal 6 F O Décl. gr. 7				60 Emplois		
Nominal 3	P C S Cardinal 1 Ordinal 2 B K Distributif 3 Multiplic. 4 Adv. ord. 5 E N Adv. mult. 6	S P Nominatif A J Vocatif B K Accusatif C L Génitif D M Datif E N Ablatif F O Locatif G P Indéclinable Z				Emploi sub. 1 adj. 2 adv. 6	
Adjectif-pronom 4	Personnel 1 Possesif 2 Réfléchi 3 Possesif réfléchi 4 Démonstratif 5 Relatif 6 Interrogatif 7 Indéfini 8		Mode	Temps	du verbe subordonné régi par un pronom relatif ou interrogatif (cf. verbe)		
Verbe 5	A P D S-D 1e conj. 1 A J 2e conj. 2 B K S 3e conj. 3 C L T 4e conj. 4 D M 4e bis 5 E N Anomal 6 F O	S P 1e A J 2e B K 3e C L	Indicatif 1 Impératif 2 Subjonctif 3 Participle 4 Adj. verbal 5 Gérondif 6 Infinitif 7 Supin en -UM 8 Supin en -U 9	Présent 1 Imparfait 2 Futur simple 3 Parfait 4 Plus-que-parfait 5 Futur antérieur 6 -us fui/fuerim/fuisse 7 -us fueram/fuissem 8 -us fuero/urus fuissis 9	61 Genre - Trème		
Adverbe 6	Relatif 6 Interrogatif 7 Négligé 8 Interrogatif-négligé 9 Comparatif 12 Superlatif 11		-Mode	Temps	du verbe subordonné régi par un adverbe relatif ou interrogatif (cf. verbe)		
Préposition 7	Type MECUM 1	Cas régi 3, 4 ou 6			a) Genre commun 1 féminin 2 masc. et fémin. 3 masculin 4 masc. et neutre 5 neutre 6 b) Trème 12		
Conjonction 8	Coordination 1 Subordination 2		. Mode	Temps	du verbe subordonné régi par une conjonction de subordination (cf. verbal)		
Interjection 9							

CODE ALPHABETIQUE			
	12	11	Zéro
1	A	J	/
2	B	K	8
3	C	L	T
4	D	M	U
5	E	N	V
6	F	O	W
7	G	P	X
8	H	Q	Y
9	I	R	Z

La lettre O sera notée Ø
Le zéro sera noté 0
La conjonction du digit 1 et du zoning zéro est représentée non par une lettre mais par le signe /

PATER | 52 PERE

A	LEMME	S-L	FORME	REF	REFERENCE	ANALYSE	P	N° D'ORDRE
0				0				0
1			1 A J / 1	1	CHAP	1		1 (DEJURE) (INDIC)
2			2 B K S Z	2	PAR/VERS	2		2
3			3 C L T 3	3	N°	3		3
4			4 D M U 4	4	D'ORDRE	4		4
5			5 E N P S	5	(PAR) (PHR)	5		5
6			6 F O W 6	6		6		6
7			7 G X 7	7		7		7
8			8 H Q Y 8	8				
9			9 I R Z 9	9				

LEMME S-L

PATERNEL | 055

MERE | 055

PERE | 055

A	LEMME	S-L	FORME	REF	REFERENCE	ANALYSE	P	N° D'ORDRE
0				0				0
1			1 A J / 1	1	CHAP	1		1 (DEJURE) (INDIC)
2			2 B K S Z	2	PAR/VERS	2		2
3			3 C L T 3	3	N°	3		3
4			4 D M U 4	4	D'ORDRE	4		4
5			5 E N V S	5	(PAR) (PHR)	5		5
6			6 F O W 6	6		6		6
7			7 G P X 7	7		7		7
8			8 H Q Y 8	8				
9			9 I R Z 9	9				

LEMME S-L

PATER | 52 PERE | 055

A	LEMME	S-L	FORME	REF	REFERENCE	ANALYSE	P	N° D'ORDRE
0				0				0
1			1 A J / 1	1	CHAP	1		1 (DEJURE) (INDIC)
2			2 B K S Z	2	PAR/VERS	2		2
3			3 C L T 3	3	N°	3		3
4			4 D M U 4	4	D'ORDRE	4		4
5			5 E N P S	5	(PAR) (PHR)	5		5
6			6 F O W 6	6		6		6
7			7 G X 7	7		7		7
8			8 H Q Y 8	8				
9			9 I R Z 9	9				

LEMME S-L

LISTE DES MOTS DANS L'ORDRE DE FREQUENCE DECROISSANTE

SVM 1	5476	NVLLVS	323.	ANIMAL	177
QVI 1	3587	CVM 2	321	NVNC	177
ET 2	3553	DE	318	QVAM 2	177
NON	2634	VIRTVS	315	VT 1	174
IN	2173	QVIDEM	304	TANTVM 2	171
HIC 1	1979	VNVS	303	CAVSA	170
ILLE	1712	SCIO	302	LOCVS	170
SI	1045	INTER	300	TEMPVS 1	168
TV	1022	IDEM	284	DIES	167
QVIS 1	1008	CORPVS	283	PORTVNA	165
SED	999	SINE	282	ALTER	162
AD	980	INQVIO	275	VITIVM	160
POSSVM 1	838	MALVM 1	273	TIMEO	155
OMNIS	827	NATVRA	272	FERO	154
SVI 1	811	QVIDAM	272	BEATVS	148
EGO	790	SIC	266	VOLVPTAS	147
IS	772	QVOQVE	264	MANVS 1	146
VT 4	742	NISI	263	DEINDE	145
FACIO	702	PER	258	NE 3	143
DICO 2	682	MORS	256	PHILOSOPHIA	142
NEC 2	648	AGO	247	VERBVM	142
EX	646	DO	247	PONO	141
ENIM 2	645	VIVO	243	IVDICO	134
QVAM 1	642	AN	242	PROSVM 1	134
HABEO	638	FIO	240	INVENIO	132
QVE	632	ITAQVE 2	239	SEQVOR	132
ANIMVS	612	NOSTER	239	DOLOR	131
ALIVS	568	ET 1	233	PARVM 2	131
IPSE	567	QVIA	233	QVIS 2	130
CVM 3	566	TAM	228	CREDO	128
RES	566	DEBEO	223	VIS	128
BONVM	559	ETIAM	223	NE 4	124
SVM 2	543	QVAERO	222	RATIO	121
NOS	539	MAGIS 2	221	SAPIENTIA	121
ALIQVIS	537	PARS	218	AMICVS 1	120
SVVS	528	QVOD 2	217	BENE	118
AVT	525	PVTO	207	LICET 1	117
MAGNVS	505	QVISQVIS 1	207	QVOMODO 2	116
ISTE	492	NEMO	202	ATQVE 1	115
AB	479	TAMEN 2	198	MODVS	114
VIDEO	468	TVVS	198	AIO	113
NIHIL	447	MEVS	192	MORIOR	113
MVLTVS	418	IAM	190	COGITO	112
BONVS	413	DEVS	188	AVDIO	109
AC 1	385	SAPIENS 1	188	TAMQVAM 1	109
VITA	373	VENIO	188	SVPERVS	108
VOLO 3	373	NAM 2	185	TVNC	108
ERGO 2	337	TOTVS	185	GENVS 1	107
AVTEM	334	VIR	185	MODO 1	107
HOMO	323	ITA	182	APVD	106

ECHANTILLON DE CHAMPS NOTIONNELS

4	PURETE INNOCENT SACRE VENERER PUDEUR PURIFIANT	CHASTETE PUR SAINT VIERGE VIRGINAL PURIFICATION	CORROMPRE PURIFIER VENERABLE CONSACRE VIRGINITE SACRILEGE
49	EPOUSE MARIAGE PRETENDANT-MARIAGE CONCUBINE GENDRE MARIE	EPOUSER MARIEE-NON RECHERCHER-MARIAGE CONJUGAL MARI	EPOUX NUPTIAL CELIBATAIRE DIVORCE ADULTERE
50	ERRANT EXILER	ERRER EXIL	EXILE
51	DESESPEREMENT INESPERE SOUHAITER VOTIF	ESPERER DESESPERER SOUHAIT	ESPOIR DESESPoir VOEU
52	ENSEVELIR MORTEL PERIR VIE FUNEBRE VIVANT SURVIVANT	EXPIRER MOURIR PERIR-FAIRE VIVRE FUNERAILLES FUNESTE	MORT MOURIR-MALHEUREUX TOMBEAU ENTERRER SURVIVRE DEUIL
53	ACTION FABRIQUER TRAVAIL	AFFAIRE FAIRE TRAVAILLER	AGIR FAIRE-A
54	FEMININ FILLETTE MASCULIN	FEMME GARCON	FEMME-JEUNE HOMME
55	FAMILLE FRERE PARENT SOEUR MARATRE	FILS MERE PATERNEL AIEUL MATERNEL	FILLE ORPHELIN PERE FRATERNEL

047	ENTRER			4
	SORTIE	#	55	4
	SORTIR	#	55	4
	SORTIR-FAIRE	#	55	4
049	EPOUSE	#	55	4
	EPOUSER	#	55	4
	GENDRE	#	55	4
	MARI	*	55	
	MARIAGE	*	55	
	MARIAGE-DU			4
	PRETENDANT	*	55	
050	ERRANT			4
	ERRER	#	55	4
	EXIL	*	55	
	EXILE	*	55	
051	ESPOIR	*	55	
	VOEU			4
052	ENTERRER	*		
	FUNERAILLES	*	55	
	FUNESTE	*	55	
	MORT	#	55	58
	MORTEL			58
	MOURIR	#	55	58
	PERIR	#	55	58
	TOMBEAU	*	55	
	VIE	#	55	58
	VIVANT	#	55	58
	VIVRE	*	55	
053	ACTION	#	58	4
	FAIRE	#	58	4
	TRAVAIL			4
	TRAVAILLER-FAIRE			4
054	FEMININ	*	37	
	FEMME	#	37	58
	HOMME	*	37	
	MASCULIN			58
055	AIEUL	*	55	
	FILLE	#	55	4
	FILS	#	55	4
	FRATERNEL	#	55	4
	FRERE	#	55	4
	MARATRE	*	55	
	MERE	#	55	4
	PATERNEL	#	55	4
	PERE	#	55	4
	PERE-BEAU	#	55	4
	SOEUR	*	55	

	1 ARME	4 PURETE	37 DIVIN	55 FAMILLE	58 VIOLENCE	61 AMOUR - HAINE
1 0 0		018 122		009 043 063 066	099	
4 0 0 0	019 021 046 070	042	026	038 067 065 067 088 093	008 020 111 124	010
37 0		113	059	108	054	
55 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	039 064 072 075 087 091	001 003 004 005 006 013 017 040 047 049 050 051 052 056 058 069 071 073 078 082 083 086 090 100 106 109 115 117 119	045 074 084 097 103 105	126	011 028 029 034 052 060 076 077 079 080 094 095 101 107 120 125	025 068 083 102 114 123
58 0 0 0 0	024	030 035 053	012 022 023 027 046	002 007 016 032 037 081 098 110 112 118	089	
61 0 0				014 041 061 116	036 121	

0
0
0

INCLUSIONS DANS LE CONTEXTE 004 PURETE

CHAMP NOTIONNEL 055 FAMILLE

	00018	LEMME(S)	00099	OCCURRENCE(S)
FRATER			00004	
FRATERNVS			00001	
GENITOR			00001	
MATER			00004	
NATA			00001	
NATVS	1		00003	
PARENS			00001	
PATER			00004	
PATERNVS			00001	
SOCER			00001	

	INCLUSION	00010	LEMME(S)	00021	OCCURRENCE(S)
--	-----------	-------	----------	-------	---------------

Annexe 7