



**HAL**  
open science

## Perséides

Étienne Brunet

► **To cite this version:**

Étienne Brunet. Perséides. Postérités de Sain-John Perse, E. Caduc, May 2000, Nice, France. pp.329-349. hal-01571928

**HAL Id: hal-01571928**

**<https://hal.science/hal-01571928>**

Submitted on 4 Aug 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## PERSÉIDES

Etienne Brunet

La constellation des chercheurs qui entoure Saint-John Perse forme un cercle étroit et savant qui admet pourtant les visiteurs de passage, ces étoiles filantes que, dans cette région polaire du ciel poétique, on nomme perséides. Comme nous venons d'une constellation voisine qui gravite autour de Giraudoux, on autorisera peut-être les flashes intempestifs du touriste. Giraudoux et Saint-John Perse sont de la même génération (Giraudoux est l'aîné); ils ont embrassé, un peu par hasard et sur le tard, la carrière consulaire et diplomatique, après quelques essais littéraires qui leur ont valu, à l'un comme à l'autre, la protection de Philippe Berthelot et le patronage de Claudel, l'illustre devancier. Nés de la même couvée, ils ont dû se croiser bien des fois, dans les ambassades étrangères ou dans les couloirs du Ministère. Dès leur entrée en fonction, qui coïncide avec la guerre, ils ont des missions semblables, l'un en Amérique, l'autre au Japon, et au terme de leur carrière, qui coïncide avec la seconde guerre mondiale, ils ont des postes de responsabilité dans le même gouvernement, l'un comme Commissaire à l'Information, l'autre comme Secrétaire général au Quai d'Orsay. Et pourtant nous n'avons guère trouvé le nom de Saint-John Perse (non plus que d'Alexis Leger) sous la plume de Giraudoux (il est vrai que nous n'avons pas fouillé les rapports consulaires). Et semblablement le nom de Giraudoux n'apparaît guère dans la correspondance de Saint-John Perse et nous ne l'avons rencontré qu'une seule fois dans le volume de la Pléiade, au détour d'une lettre adressée à Claudel: "Je n'ai pas revu Berthelot, pour ne le point importuner, mais j'ai su par Giraudoux que votre *Protée* avait fait son entrée en France dans la valise diplomatique."<sup>1</sup> Mais cette preuve de leur relation n'est pas celle de leur amitié et on a tout lieu de penser que les deux diplomates-écrivains ont été plutôt des rivaux que des camarades et qu'ils ont même cultivé à plaisir leur différence: Giraudoux, suivant en cela l'exemple de Claudel, n'a jamais caché son statut parallèle d'écrivain, et il lui arrivait de rédiger ou de corriger quelque manuscrit dans l'exercice de ses

---

<sup>1</sup> Lettre à Paul Claudel, Consul général de France à Hambourg, Paris, janvier 1914, la *Pléiade*, p. 726.

fonctions consulaires, tandis que Alexis Léger a voulu séparer radicalement l'écrivain et le haut fonctionnaire et s'est abstenu de publier en France durant son mandat. Au dilettantisme de l'un s'oppose le professionnalisme de l'autre. Et même en littérature leur destin diverge. Cette fois c'est Saint-John Perse qui suit l'exemple de Claudel et s'établit sur le haut plateau poétique, tandis que Giraudoux reste dans la vallée (il n'a écrit que des vers de mirliton) et papillonne dans les genres à succès, romans, essais et théâtre.

On pourrait approfondir le parallèle mais notre objet est plutôt d'appliquer à Saint-John Perse le traitement électronique auquel nous avons soumis Giraudoux et quelques autres écrivains. Nous ne sommes pas certain d'avoir l'aval de Saint-John Perse pour une intrusion qui ressemble à une effraction. Mais tant d'écrivains ont été approchés, sans trop de respect, par l'ordinateur, de la Bible à Shakespeare et de Balzac à Proust, que l'autorisation expresse de l'auteur n'est plus nécessaire. Au reste les travaux de dépouillement ont été entrepris officiellement au sein même de la Fondation Saint-John Perse, par l'un de ses membres, Éveline Caduc, qui a enregistré sur un support électronique l'oeuvre poétique de Saint-John Perse, telle qu'on la trouve dans l'édition de la Pléiade, soit neuf recueils répartis comme suit:

*Tableau 1. Composition du corpus Saint-John Perse*

N°	TITRE	OCCURRENCES	VOC&LES	Prob P	Prob Q	ABREGÉ	CODE
1	Éloges	6550	1723	.0746	.9254	Éloges	Él
2	Rois	3196	944	.0364	.9636	Rois	Ro
3	Anabase	6518	1631	.0743	.9257	Anabase	An
4	Exil	10366	2245	.1181	.8819	Exil	Ex
5	Vents	18991	3682	.2164	.7836	Vents	Ve
6	Amer	30448	4666	.3469	.6531	Amer	Am
7	Chroniques	3930	1112	.0448	.9552	Chroniques	Ch
8	Oiseaux	4464	1297	.0509	.9491	Oiseaux	Oi
9	Équinoxe	3306	1029	.0377	.9623	Équinoxe	Éq
	TOTAL	87769	9843				

Ces données textuelles ont été indexées, triées, comptées, référencées, bref soumises à divers traitements qui conduisent à l'hypertexte, c'est à dire à un état du texte où chaque mot est accessible en accès direct dans son environnement immédiat. Aux fonctions documentaires réparties sur la marge haute et horizontale de la figure 2, s'ajoutent les fonctions statistiques, échelonnées sur la marge verticale, à droite.

Figure 2. Le menu principal de l'hypertexte Saint-John Perse

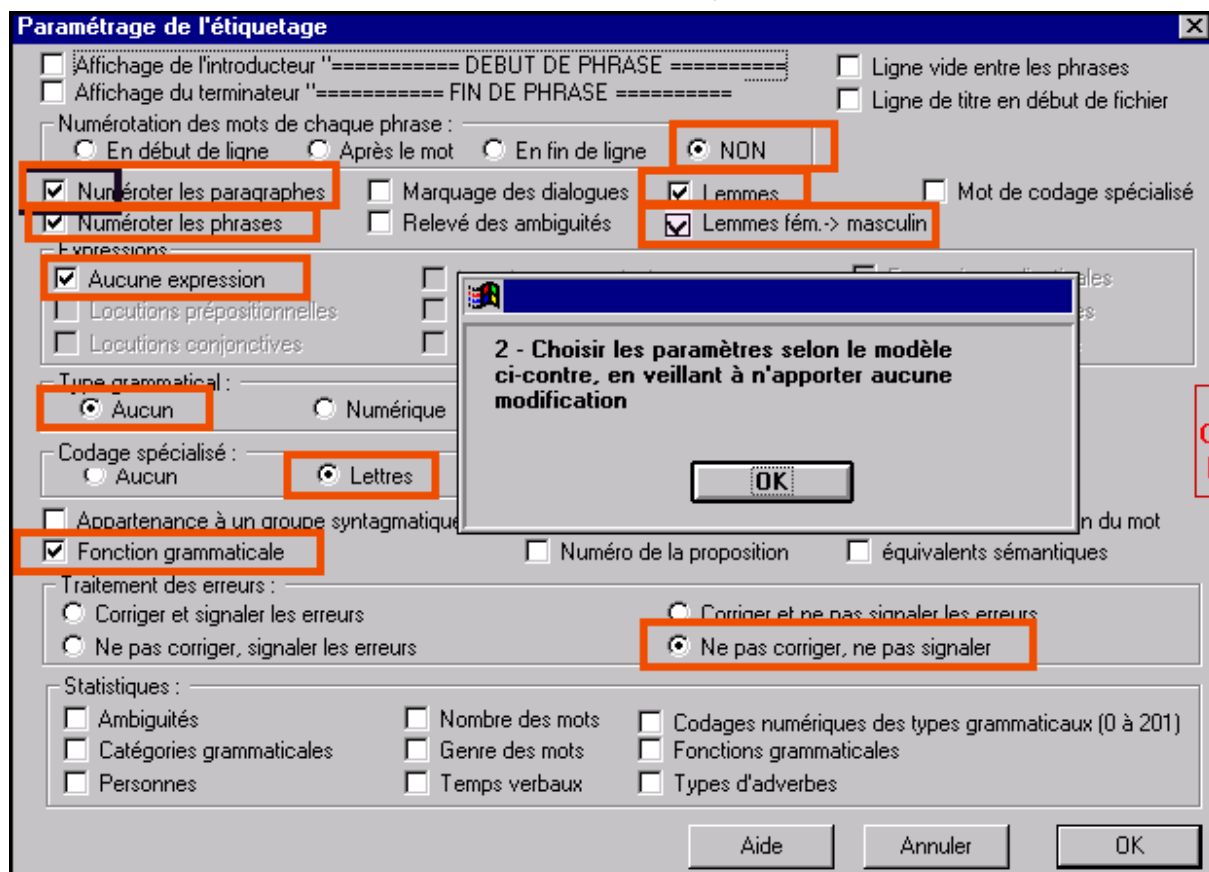


Cette base existe depuis longtemps, comme beaucoup d'autres consacrées à des monographies d'écrivains<sup>2</sup>. Si nous la présentons aujourd'hui, c'est parce qu'elle a bénéficié récemment de nouveaux développements, relatifs à la lemmatisation. Sous ce nom barbare, inconnu du *Robert*, se cache un traitement linguistique qui désambiguïse les homographes, si nombreux dans un texte français. L'exemple le plus fréquent est celui de l'article défini, dont la graphie se confond avec celle du pronom personnel de la troisième personne. Que dire de *le, l', la* et *les*, si la séparation des emplois n'est pas faite? Et que dire de *que*, si la distinction de la conjonction et du pronom relatif n'est pas établie? Parmi les mots-pleins la confusion n'est pas moins redoutable: *marche* et *marcher* partagent des formes communes (et aussi des sèmes communs), mais les

<sup>2</sup> Le même logiciel *Hyperbase*, créé par nos soins, a été appliqué à de nombreuses intégrales (ou quasi intégrales): Rabelais, Montaigne, Corneille, Molière, La Fontaine, Racine, Pascal, Marivaux, Voltaire (partiellement), Rousseau, Chateaubriand, Balzac, Hugo (partiel), Baudelaire, Rimbaud, Sand, Flaubert, Maupassant, Verne, Proust, Eluard, Breton, Mammeri, Chraïbi, Gracq, Le Clézio, etc.

graphies homographes se rattachent à des catégories concurrentes, entre lesquelles seul le texte permet de trancher. L'analyse est simple si l'on rencontre *il marche* ou *la marche*. Mais dans beaucoup de cas la recherche doit s'exercer plus en amont ou plus en aval. On utilisera pour ce faire un logiciel du commerce, *Cordial*, qui a pour fonction première de corriger les textes. Mais les concepteurs de ce produit sont ouverts à la recherche universitaire et une version particulière du logiciel est destinée aux laboratoires spécialisés dans le traitement automatique de la langue, auxquels elle fournit un outil d'analyse et non plus seulement de correction. Cette version, anciennement dénommée « Cordial Université », est maintenant distribuée sous l'étiquette ANALYSEUR et correspond à la version 7 (ou 8) du produit standard. Ce programme étant automatique n'est pas exempt d'erreurs. Mais il échappe à la fatigue, à l'inconstance, à la subjectivité et finalement au renoncement qui accompagnent les entreprises de désambiguïsation manuelle. On veillera à maintenir constant le choix des paramètres proposés selon les options qui sont cochées dans l'illustration ci-dessous.

Figure 3. L'application Analyseur de Cordial



Outre le code grammatical, *Cordial* ajoute de nombreux renseignements relatifs au traitement des expressions, à la fonction dans la phrase, à la place hiérarchique du mot dans l'arbre syntaxique, et même à la classe sémantique à laquelle le mot se rattache. Nous n'avons retenu que ce qui était strictement nécessaire à l'analyse, soit la moitié des possibilités offertes dans la figure 3, à savoir: le numéro du paragraphe, le numéro de la phrase, la forme, le lemme, le code grammatical détaillé (codegram) et la fonction.

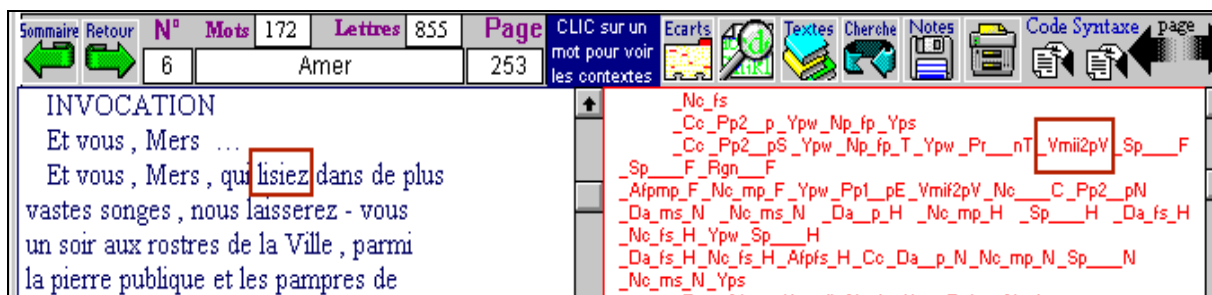
*Hyperbase* prend alors en compte les principaux éléments d'un tel fichier et les distribue séquentiellement dans quatre champs parallèles, voués respectivement aux formes, aux lemmes, aux codes et aux structures. La figure 4 met en correspondance les formes et les lemmes d'une même page de Saint-John Perse. On notera que les lemmes, dans la partie droite de l'écran, sont pourvus d'un indice numérique, afin de séparer les uns des autres les homographes. Ainsi *le 7* (dans *la grande chose*) distingue l'article du pronom codé 5 (dans *le furent*). Ces codes simplifiés qui reproduisent la classification de Muller et de Labbé (1 verbe, 2 substantif, 3 adjectif, 4 numéral, 5 pronom, 6 adverbe, 7 déterminant, 8 conjonction, 9 préposition) n'appartiennent pas en propre à *Cordial*, mais ont été dérivés de l'analyse complète fournie par *Cordial*.

Figure 4. L'alignement forme-lemme

The screenshot displays the Cordial software interface. At the top, a menu bar includes options like 'Sommaire', 'Retour', 'N°', 'Mots', 'Lettres', 'Page', 'CLIC sur un mot pour voir les contextes', 'Ecart', 'Textes', 'Cherche', 'Notes', 'Code Syntaxe', and 'page'. Below the menu, a status bar shows '6' for 'Mots', 'Amer' for 'Lettres', and '253' for 'Page'. The main window is divided into two vertical panes. The left pane contains the text of 'INVOCATION' with several words highlighted in pink boxes: 'le', 'la', and 'le'. The right pane shows the corresponding lemmatized text with numerical codes (1-9) indicating grammatical classes. A legend at the bottom right of the right pane states: 'code 5 = pronom' and 'code 7 = déterminant'.

Cette analyse complète est rendue visible, quoique peu lisible, pour peu qu'on sollicite le bouton *CODE* situé à droite de la barre de menu. Là aussi l'alignement est rigoureux, en sorte que l'on sait précisément à quel mot correspond telle ou telle analyse. Ces quatre champs sont sensibles au clic de la souris: tout objet que l'on désigne, qu'il s'agisse d'une forme, d'un lemme, d'un code ou d'une structure, renvoie aux autres occurrences où le même objet est rencontré, les relations hypertextuelles s'appliquant aux quatre champs. Mais ces relations lient aussi ces champs entre eux, en sorte qu'en cliquant sur un code grammatical dans le champ de droite (par exemple *\_Vmii2pV*, soit *Verbe, non auxiliaire, indicatif, imparfait, 2e personne du pluriel, base de proposition*), on obtient successivement en vidéo inverse tous les verbes qui répondent strictement à ce codage.

Figure 5. L'alignement forme-code



L'indexation et toutes les opérations subséquentes sont alors répétées quatre fois, au niveau des structures syntaxiques (c'est à dire des séquences ordonnées des parties du discours), puis des codes grammaticaux, puis des lemmes, puis des formes.

À l'issue de ce traitement, on obtient quatre index (figure 6) qui réagissent pareillement au clic de la souris. La forme (ou le lemme ou le code ou la structure) qu'on désigne montre le détail de ses occurrences, parmi lesquelles l'utilisateur fait son choix pour se référer au texte.

Figure 6. Les quatre index issus de Cordial

Formes	Lemmes	Codes
1 écaillère , 6 1	1 déchet 2 , 7 1	64 _afpmp_d , 1 3 2 1 3 1 4 1 4
1 écaillères , 6 1	1 déchiffrer 1 , 6 1	5 1 4 6 2 7 7 1 9 3
9 écailles , 1 1 4 1 5 2 6 5	1 déchiffreur 2 , 5 1	71 _afpmp_f , 1 6 2 2 3 4 4 6 5
1 écale , 7 1	6 déchirement 2 , 6 4 7 1	20 6 2 2 7 5 8 3 9 3
2 écarlate , 6 1 9 1	8 1	3 _afpmp_g , 6 3
3 écarlates , 1 1 6 2	3 déchirer 1 , 6 3	
7 écart , 2 1 5 3 6 3	2 d	
1 écartant , 3 1	2 d	
1 ecchymose , 6 1	1 d	
1 échange , 5 1	1 d	
1 échangées , 6 1	1 d	
1 échangèrent , 7 1	4 d	
1 échangés , 3 1	2 d	
2 échappe , 5 1 8 1	1 décoration 2 , 6 1	
1 échappent , 3 1	1 découper 1 , 1 1	
2 écharde , 5 1 6 1	1 découpler 1 , 6 1	
1 échauffait , 6 1	1 découpure 2 , 8 1	
1 échauffe , 6 1	2 découvert 2 , 2 1 6 1	
1 échauffées , 6 1	1 découverte 3 , 7 1	
3 échéance , 3 1 4 1 5 1	1 découvrir 1 , 1 1	
1 échelle , 4 1	1 décret 2 , 6 1	
1 Échelles , 8 1	2 décroître 1 , 5 1 6 1	
1 évechevèle , 3 1	1 dédaigneuse 3 , 5 1	
2 évechevement , 5 1 6 1	1 dédale 2 , 5 1	
1 échine , 5 1	1 dedans 6 , 1 1	

Adjectif, qualificatif, positif, masculin, pluriel, objet indirect.

OK

Structures

1 lvrncvdsn , 7 1

1 lvrndrna , 6 1

1 lvrndsdn , 4 1

1 lvrnsdn , 6 1

1 lvrnsdcdn , 6 1

1 lvrnscdn , 4 1

1 lvrnsdncvdsn , 5 1

Les sous-  
fréquences dans les  
textes sont  
réparties en  
binomes, dont le  
premier élément  
indique le numéro  
d'ordre du texte ,  
le second la  
fréquence du mot

Cliquez sur un  
mot pour l'activer

← →

Index hiérarchique

Chercher un mot

Graphique

Éditer

S'il s'agit d'un code, dont la signification peut être opaque, le décryptage est assuré et traduit en clair, comme dans l'exemple ci-dessous, relatif au code *\_afpmp\_d*, c'est à dire à un adjectif qualificatif, positif, masculin pluriel, dans un groupe objet indirect. Lorsqu'on veut, non plus reconnaître un code particulier, mais rassembler les mots qui ont reçu le même codage (par exemple pour constituer une concordance ou un histogramme), on est renvoyé à une page spéciale (figure 7) qui dénombre toutes les combinaisons possibles. Car *Cordial* pousse loin l'analyse, en relevant pour chaque mot la catégorie, la sous-catégorie, le genre, le nombre, la fonction et s'il s'agit d'un verbe le temps, le mode et la personne. Un clic dans une option provoque alternativement l'activation ou la désactivation correspondante. Certaines options sont impliquées ou exclues automatiquement, dès qu'une autre est choisie, de telle façon qu'il y ait toujours cohérence. Car il serait absurde de sélectionner le futur d'un substantif ou le féminin d'un verbe à l'infinitif. Chaque clic modifie le filtre dont l'affichage apparaît dans une fenêtre, en haut et à droite de l'écran, avec sa traduction en clair. Toute colonne non intéressée par la sélection est remplie par défaut par un joker, dont l'effet est d'admettre tout code qu'on rencontre à cet endroit. Ainsi dans l'exemple choisi la colonne 3 n'ayant pas été



sélectionnée dans le code *Af\_ms*\_\_, tous les adjectifs seront retenus, quel que soit le degré, positif ou comparatif. De même le vide rencontré dans la colonne 7 laissera la sélection indifférente à la fonction dans la phrase.

Une fois que la sélection est faite, elle est transmise à la fonction appelante, qui délivre un contexte, une concordance, ou une liste, c'est à dire un tableau à deux dimensions dont chaque ligne dresse le profil d'une sélection grammaticale à travers le corpus et chaque colonne celui d'un texte du corpus à travers les codes grammaticaux.

Figure 7. Le choix d'un code grammatical

Catégorie 1	Sous-cat.2	Mode 3	Temps 4	Personne 5	Code choisi	1	2	3	4	5	6	7	Retour	Sommaire
Verbe V	principal m auxiliaire a	Infinitif n Indicatif i Subjonctif s Conditionnel c Impératif f Participe p	Présent p Imparfait i Passé s Futur f Subjonctif présent r Subjonctif imparfait m Participe passé a	1re pers. 1 2e pers. 2 3e pers. 3	<b>Af_ms</b>									
Substantif N	nom commun c nom propre p													
Adjectif A	qualificatif f ordinal o	Positif p Comparatif c	Genre 4 Masculin m Féminin f	Numéral Mc Interjection Ij										
Déterminant D	article a démonstratif d interrogatif i indéfini t		Nombre 5-6 Singulier s Pluriel p	Préposition Sp Adverbe R										
Pronom P	personnel réfléchi x pers. non réfléchi p possessif s démonstratif d interrogatif t indéfini i relatif r		Fonction 6 sujet n objet direct a objet indirect d	Conjonction C Ponctuation Y										
Effacer Continuer Cliquer sur les critères souhaités puis sur le bouton CONTINUER													Fonction 7 A - attribut du sujet B - groupe attribut du sujet C - objet direct D - groupe objet direct E - objet indirect F - groupe objet indirect G - complément d'agent H - circonstanciel K - circ. de temps L - circ. de lieu M - apposition N - groupe apposition O - apostrophe P - groupe apostrophe Q - compl. de négation S - sujet T - groupe sujet U - pronominalisation V - base de proposition Y - sujet réel Z - groupe sujet réel 1 - ajout à l'adjectif 2 - reprise du COD 3 - reprise du COI 4 - reprise du circonst. 5 - ajout au nom 6 - ajout au pronom 7 - reprise du sujet 8 - ajout au verbe	
Choisir la combinaison souhaitée. Un clic sur une option sert alternativement à activer ou désactiver la sélection. Les options inscrites dans la zone bleue sont réservées aux verbes. Certaines autres aux adjectifs ou aux pronoms. Les options 4 (genres), 5-6 (nombre) et 7 (fonction) concernent toutes les parties du discours, sauf les invariables. Le programme interdit les choix incohérents. Une fois réalisée la sélection, cliquer sur CONTINUER pour la transmettre au traitement en cours. (Le numéro des options indique la colonne intéressée dans le code).														

Le choix des structures syntaxiques se fait de la même façon en choisissant dans l'ordre les catégories qui forment la séquence souhaitée, par exemple déterminant + adverbe + adjectif + substantif + pronom relatif, à chaque endroit de la séquence une constante (forme, lemme ou code grammatical détaillé) pouvant se substituer à la variable. Ainsi dans la figure 8 deux contraintes sont imposées à la sélection: l'adjectif doit appartenir au lemme *beau* et le verbe qui

suit le relatif doit être au subjonctif (code `_V_s`). Si l'on ajoute que la fonction de chaque mot peut être isolée dans le code, et qu'un balisage sémantique est appliqué aussi au texte, on mesure combien l'étude, traditionnelle ou statistique, du texte s'enrichit de mille possibilités.

Figure 8. Choix d'une structure syntaxique

STRUCTURE SYNTAXIQUE		Retour	Sommaire
		1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0	
<i>Valeurs</i>	<i>Structure choisie</i>	<b>dranlv</b>	
1	<input type="text" value="d"/>	Déterminant Adverbe BEAU 3 Substantif Pronom_relatif_V_S	
2	<input type="text" value="r"/>		
3	<input type="text" value="beau 3"/>	<b>Effacer</b>	Verbe v Adverbe r
4	<input type="text" value="n"/>	<b>Continuer</b>	Substantif n Déterminant d
5	<input type="text" value="l"/>	Choisir dans l'ordre les éléments que l'on souhaite incorporer dans la combinaison syntaxique, en cliquant sur les boutons à droite de l'écran.	
6	<input type="text" value="_v_s"/>	Lorsqu'il s'agit d'une recherche de contexte ou d'une concordance, une valeur particulière (forme, lemme ou code grammatical détaillé) peut être donnée à un ou plusieurs éléments de la structure (cliquer sur les boutons à gauche de l'écran).	
7	<input type="text"/>	Pour corriger une proposition, utiliser le bouton EFFACER..	
8	<input type="text"/>	Quand le choix est fait, cliquer sur CONTINUER.	
9	<input type="text"/>	Pronom_relatif l	Subordination b
10	<input type="text"/>	Autre_pronom p	Interjection i
11	<input type="text"/>		

Ce n'est pas le lieu d'entreprendre l'exploitation d'un tel corpus lemmatisé, que vous préférons livrer dans sa pureté virginale aux spécialistes du poète. Là où l'étude des formes permet une première approche, celle des lemmes apporte le plus souvent une confirmation. Ainsi les 9 recueils de notre corpus se répartissent de la même façon lorsqu'est mesurée la distance entre leurs vocabulaires, lemmatisés ou non. De même les listes de spécificités obtenues pour chaque texte à partir des formes et des lemmes ont beaucoup d'éléments communs, quoique les informations qu'elles donnent pour les verbes soient beaucoup plus sûres si la lemmatisation a opéré le regroupement des formes fléchies. Mais dès qu'on aborde la syntaxe et les faits de style, l'étude des formes montre vite ses limites et ses faiblesses: l'approche biaisée des mots grammaticaux ne permet qu'une approximation timide, car beaucoup de mots-outils sont homographes et servent à plusieurs usages, comme les couteaux

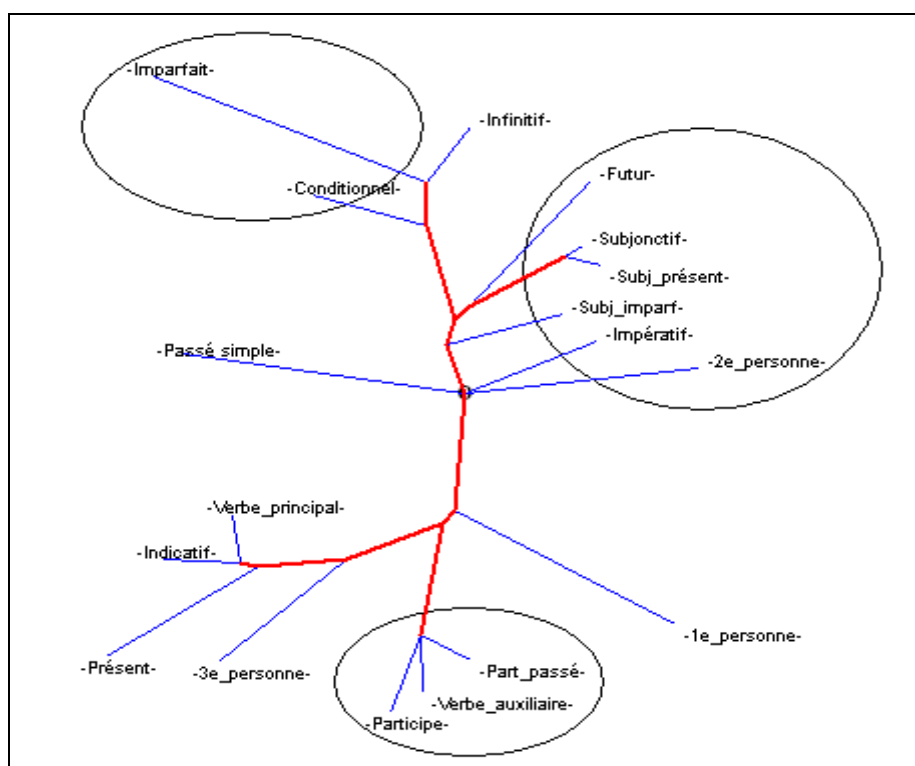
suisses. Sans lemmatisation préalable, bien des aspects du texte restent inaccessibles: le genre grammatical, le nombre, la fonction dans la phrase, le temps verbal, le mode, la personne, les parties du discours et leurs multiples combinaisons réalisées dans les syntagmes et les structures syntaxiques, des plus simples (digrammes ou trigrammes) aux plus complexes, tout cela échappe aux formes brutes.

C'est le verbe qui gagne le plus à l'étiquetage: si l'on s'intéresse à son contenu sémantique, le regroupement des formes d'un même verbe offre un avantage indéniable, et si l'on étudie un temps verbal et qu'on regroupe les verbes qui partagent le même codage et le même temps, l'avantage est encore plus net. Et faute de pouvoir aborder les points de vue qu'on vient de passer en revue, c'est le système verbal qu'on voudrait étudier dans notre corpus, à titre d'exemple. Lorsqu'il s'agit de l'emploi des verbes, on a tout lieu de penser qu'un écrivain y porte attention. Le choix qu'il fait du passé ou du présent, de la première ou de la troisième personne, a des conséquences importantes pour la conduite du discours et une telle décision ne saurait être inconsciente. Le système verbal s'étagant sur plusieurs plans: le mode, le temps et la personne (sans compter le nombre, l'aspect et d'autres paramètres), on pourrait isoler successivement les trois plans principaux (les seuls que relève *Cordial*), ou bien les croiser et, par exemple, consacrer une ligne du tableau à la troisième personne du pluriel du présent de l'indicatif des verbes auxiliaires (croisement de 5 variables). Pour une première approche il nous a paru prudent de ne retenir que les grandes divisions, sans croisement, mais sans exclusive. En étudiant ensemble, comme variables indépendantes, les modes, les temps et les personnes, on se donne le moyen de repérer lequel de ces trois paramètres est le plus discriminant, mais aussi quelle interaction s'exerce entre les uns et les autres.

Une première réponse est dans le graphique ci-dessous (figure 9). On y voit que le mode n'est pas la pierre de touche qui puisse servir à classer les textes et les styles. Les modes apparaissent moins excentriques que les temps, à l'exception du participe qui s'éloigne vers le bas en s'associant aux auxiliaires pour constituer les temps composés. La voix la plus forte appartient au temps; c'est elle qui impose sa loi au discours, en le sommant de choisir entre le présent et l'imparfait (le passé simple, peu employé en poésie, hésite entre les deux). La tension la plus intense (sur le graphe la distance la plus longue) est en effet celle qui oppose le présent (en bas) à l'imparfait (en haut). Quant au passé composé, il se rapproche du présent auquel il se rattache formellement. Mais ce qui est le

plus frappant, c'est l'entrecroisement des trois critères du verbe. Des alliances se nouent entre le futur, le subjonctif, l'impératif et la seconde personne, entre le conditionnel et l'imparfait, entre l'indicatif, le présent et la troisième personne. Certes ces réseaux d'influence sont parfois imposés par la grammaire (comme la concordance des temps) ou la logique (un impératif appelle la seconde personne). Mais ils dépendent aussi du genre littéraire. Ceux qu'on observe ici, là où prédomine la poésie, ne se retrouvent pas nécessairement dans le récit, où le rôle du verbe est plus important et plus différencié, l'évocation poétique se contentant souvent d'un présent intemporel<sup>3</sup>.

Figure 9. Analyse arborée des temps, des modes et des personnes



Les choix étant ainsi offerts, comment réagissent les textes? Ils ont souvent des goûts opposés, comme le suggère le graphique 10 qui représente conjointement les recueils *Gloire des Rois* et *Vents*. Là où le premier privilégie le présent, le second préfère l'imparfait. Le contraste se répercute à tous les points de la courbe, et le coefficient de corrélation ( $r = -0,73$ ) est largement

<sup>3</sup> Dans un roman, il arrive souvent que l'imparfait l'emporte sur le présent. Chez Saint-John Perse l'effectif de l'imparfait dépasse à peine le dixième du présent.

significatif. Si l'on envisage les autres recueils dans une analyse globale du corpus (analyse factorielle de la figure 11), on voit que le choix des *Rois* est partagé par les premiers recueils et que *Vents* est isolé sur la gauche, tandis que les derniers recueils (à l'exception des *Oiseaux*) s'orientent vers le haut du graphique. Ici comme sur d'autres plans, se manifeste l'évolution de l'écrivain.

Figure 10. Répartition des temps, modes et personnes dans deux recueils

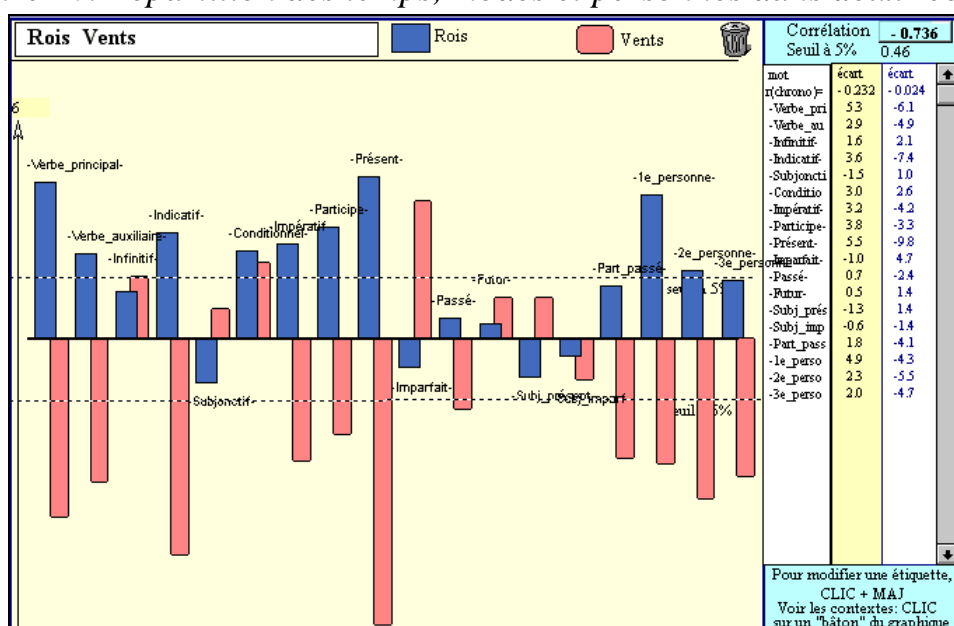
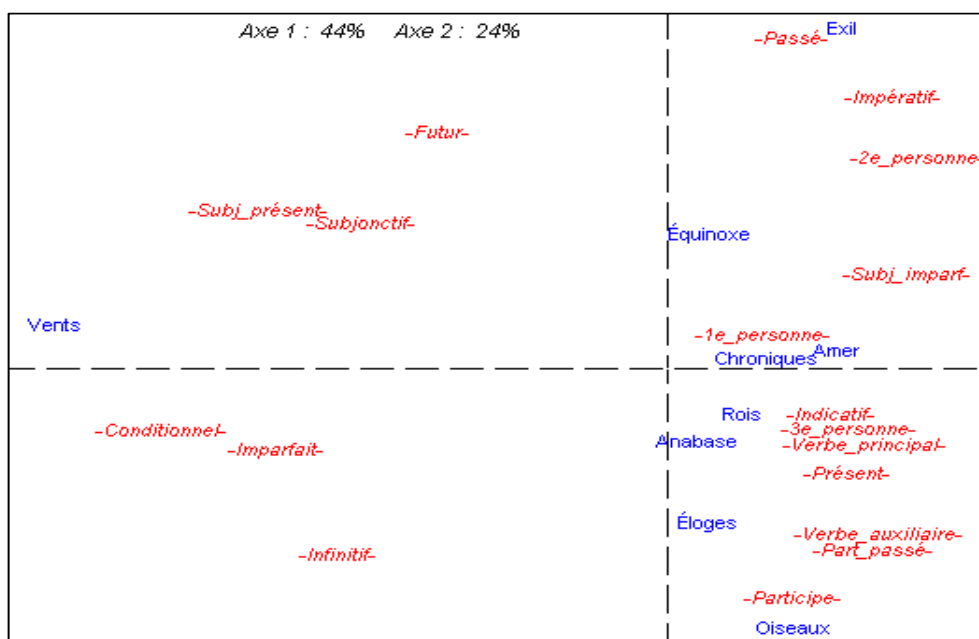
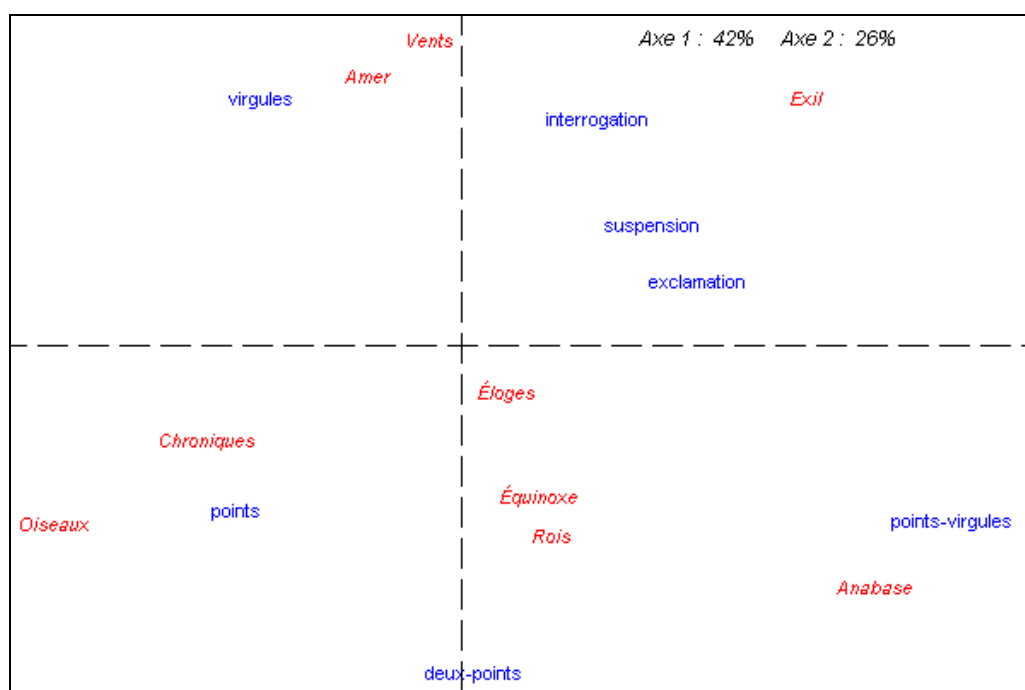


Figure 11. Analyse factorielle des verbes



Cette évolution n'est pas véritablement linéaire. On constate des retours, des mouvements courbes qui rapprochent les périodes extrêmes. Ainsi l'étude de la ponctuation - qui est ici légitime puisque l'auteur a eu le loisir de corriger les épreuves sans rien abandonner aux protes - laisse apparaître des périodes favorables aux signes affectifs (l'interrogation, l'exclamation et la suspension dans *Exil*), d'autres qui cultivent la virgule et les phrases longues (*Vents*, *Amers*), les périodes extrêmes se partageant le point (*Oiseaux*, *Chroniques*) et le point-virgule (*Éloges*, *Rois*, *Anabase*).

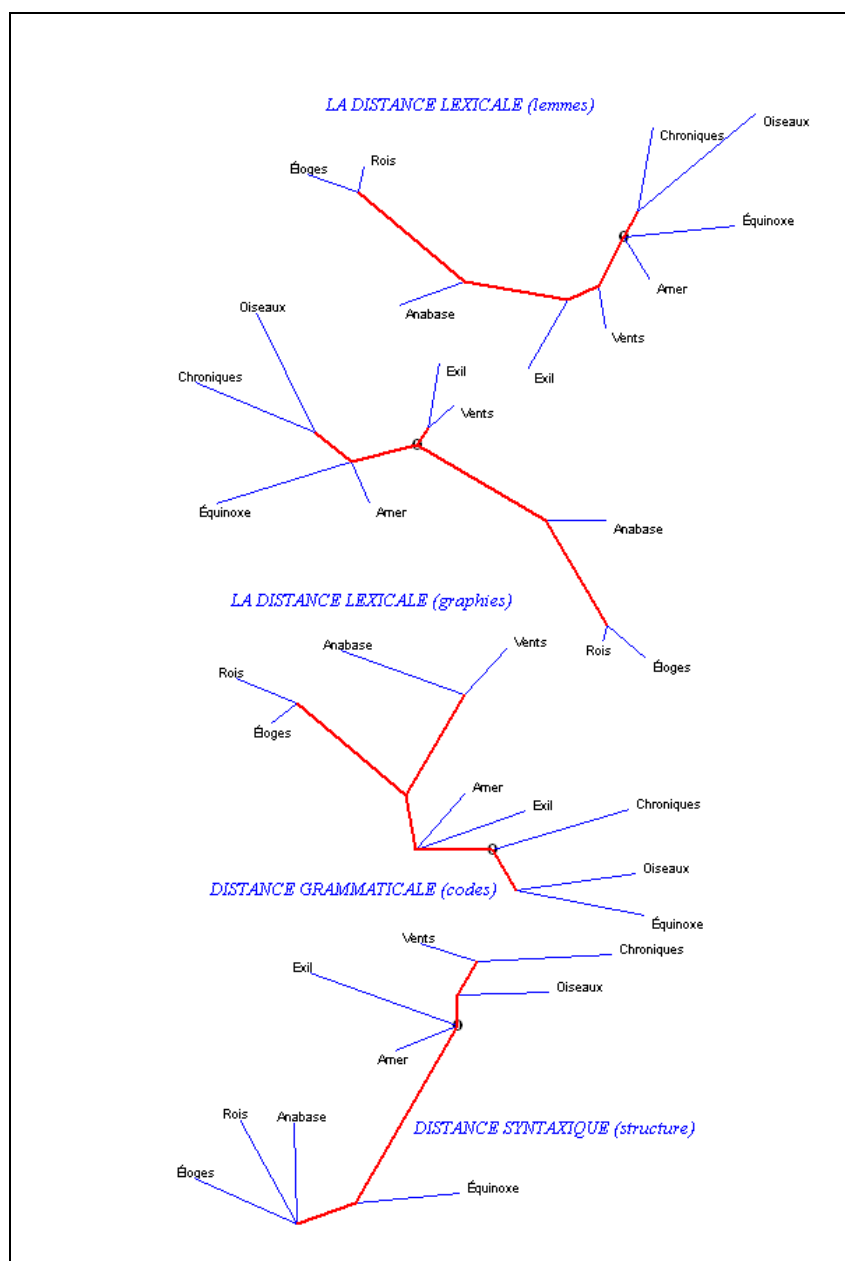
Figure 12. Analyse factorielle de la ponctuation



Suivant l'objet considéré, les variations dues au temps peuvent fluctuer. Mais si l'on considère le vocabulaire entier, la chronologie dessine une ligne stable et pure. La méthode pour calculer la distance entre deux recueils consiste à considérer tous les mots et à comparer leur répartition dans les deux textes confrontés. Si la répartition est régulière, cela rapproche les deux textes, si elle est déséquilibrée cela contribue à accentuer la distance. Mais il y a plusieurs façons de considérer les mots, selon qu'on s'intéresse aux graphies, aux lemmes, aux codes grammaticaux ou aux structures syntaxiques. Or quel que soit l'objet isolé, on obtient la convergence. La chaîne chronologique se reconnaît dans les quatre graphiques qui rendent compte de ces distances dans la figure 13. Partout

les trois premiers recueils se tiennent la main, partout les trois recueils intermédiaires occupent des positions proches, partout *Chroniques* voisine avec *Oiseaux*.

Graphique 13. La distance entre les recueils (analyse arborée)



Non content d'offrir une information grammaticale très riche, *Cordial* fournit encore des statistiques, dont certaines abordent le domaine de la sémantique et de la thématique. Un immense tableau est ainsi constitué qui comporte autant de lignes que de variables ou concepts répertoriés et autant de

colonnes que l'on compte de textes dans le corpus. Une page spéciale (reproduite ci-dessous) donne accès à ces indices et en assure la gestion, par séries homogènes, de type grammatical, rythmique ou sémantique.

Figure 14. Les 1345 indices statistiques relevés par Cordial

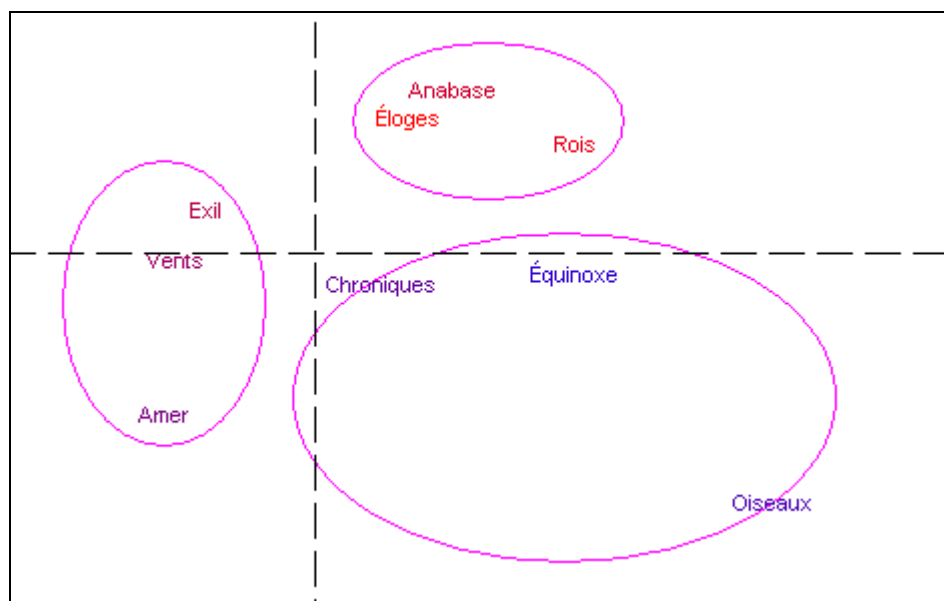
The screenshot displays the Cordial software interface. At the top, there are navigation buttons: 'Sommaire', 'Retour', 'Aide', 'CONTINUER', 'Remise à zéro', and 'CORDIAL'. A status bar indicates 'Nombre de variables: 29'. Below this, there are two main columns: 'Sélection' and 'Catalogue'. The 'Sélection' column lists various terms like 'grandes\_notions', 'ordre', 'mesure', etc. The 'Catalogue' column shows the same terms with additional labels like 'SECONDAIRE' and 'TERTIAIRE'. On the right side, there are several panels of buttons for analysis: 'TOTAUX (effectifs absolus)', 'STYLE (pourcentages)', 'NIVEAU LANGUE (coefficients)', 'USAGE (pourcentages)', 'Segmentation (pourcentages)' with buttons for 'MOYENNES 1', 'PONCTUATIONS', 'PHRASE', and 'PROPOSITIONS', 'Parties du discours (pourcentages)' with buttons for 'MOYENNES 2', 'MORPHOLOGIE', 'TYPES', 'NOMS', 'N. COMMUNS', 'N. PROPRES', 'PRONOMS', 'ADJECTIFS', 'ADVERBES', 'VERBES', 'Sémantique (coefficients)' with buttons for 'CENTRAL', 'SECONDAIRE', 'DOMAINES', 'TERTIAIRE', and 'BASES'. A text box on the right explains the selection process: 'Cliquer sur un des boutons ci-dessous, pour sélectionner les éléments qu'il contient (retouches par un clic dans les listes). Puis cliquer sur CONTINUER.'

Le résultat de l'analyse thématique confirme ce que nous avions enseigné les études qui précèdent. Les recueils là aussi forment trois groupes de trois éléments, ordonnés selon la chronologie (dans l'analyse factorielle de la figure 15). Cette robustesse et cette constance des résultats s'expliquent en partie parce que d'autres facteurs sont absents qui pourraient perturber les résultats, comme le genre littéraire. Les 9 recueils appartiennent en effet au même genre poétique. En revanche ce genre poétique est très conservateur: ni la révolution romantique, ni celle des surréalistes n'ont vraiment modifié la composition du stock lexical de la poésie, où, malgré ce qu'en dit Hugo, les mots nobles tiennent le haut du



pavé et où le cochon n'a jamais eu droit de cité. Or dans un milieu rebelle au changement, les marques du temps s'observent malgré tout.

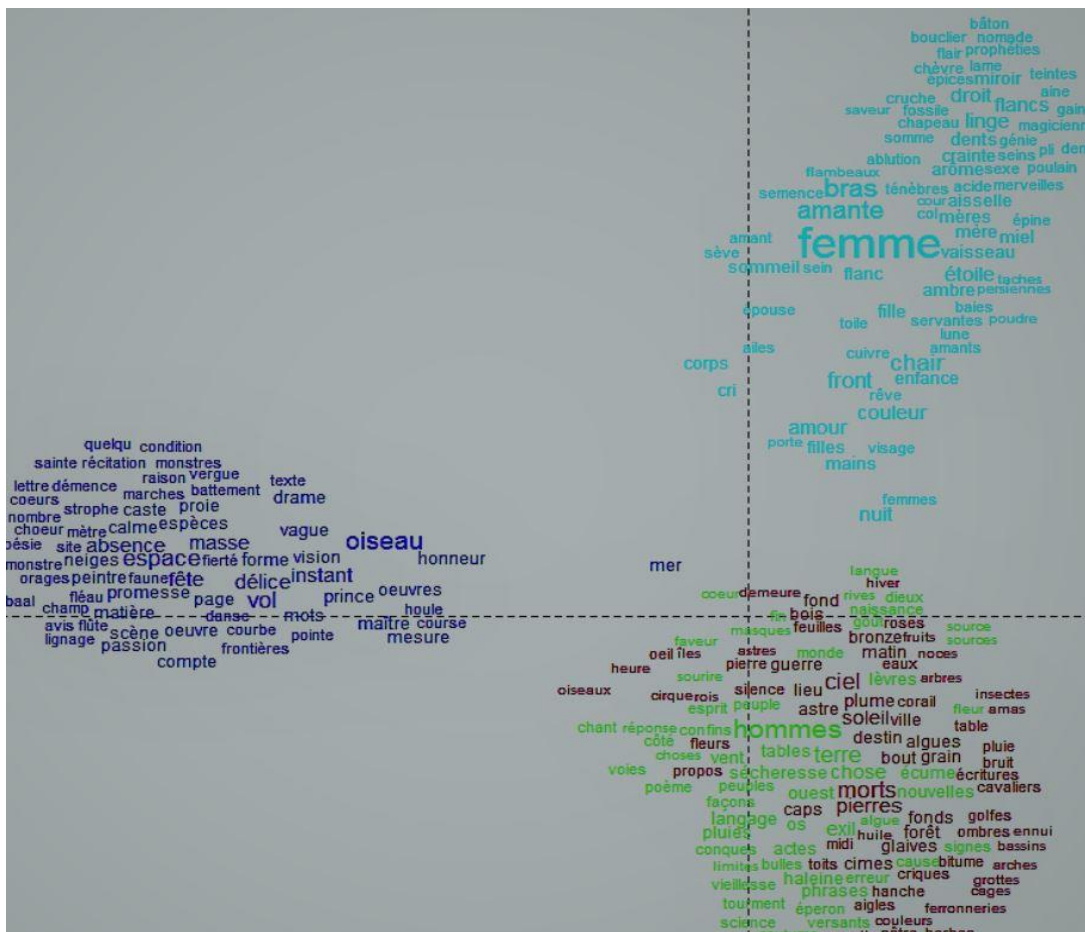
Figure 15. Analyse factorielle des thèmes relevés par Cordial



Indépendamment de *Cordial*, on a essayé une autre approche fondée sur les corrélats et empruntée au logiciel *IRAMUTEQ*. Au lieu de recevoir leur étiquette de l'extérieur, d'un thesaurus préalablement constitué, on voudrait que les mots reconnaissent d'eux-mêmes leur parenté, du seul fait de leur voisinage dans les mêmes contextes. Pour ce faire, le programme comptabilise tous les mots-pleins (au moins ceux qui ne sont pas trop rares) et enregistre toutes les rencontres, occasionnelles ou insistantes, avec les autres mots-pleins. Un lien est établi entre deux lemmes quand ils ont tendance à se rencontrer dans le même passage. L'élimination des mots-outils concourt à privilégier les relations sémantiques ou thématiques plutôt que les rapports de dépendance syntaxique. On notera enfin que la division en textes est ignorée. La cohabitation à longue distance dans un même texte n'entre pas dans le calcul. Seule compte la proximité immédiate dans la même page, là où l'on a le plus de chances de relever les isotopies.

La figure 16 (c'est encore une analyse factorielle) rend compte de ces isotopies que la couleur permet de circonscrire. La zone marquée en bleu à gauche met en relief des mots aériens comme l'*oiseau*, le *vol*, l'*espace* et la *mer*. C'est aussi là que la même analyse place les recueils *Oiseaux*, et *Rois*. Le quadrant supérieur droit est dévolu à la femme (mais aussi *mère*, *mères*, *fille*, *filles*) et se rattache plutôt aux recueils *Amers* et *Eloges*. Les mots les plus caractéristiques dans cet univers appartiennent au corps et à l'amour (*amante*, *chair*, *bras*, *front*, *mains*, *visage*, *corps*, *amant*, *amants*, *amour*). Il est plus difficile de distinguer les deux autres thèmes qui se partagent le dernier quadrant, en bas à droite. La couleur verte met en exergue le mot *hommes* et

Figure 16. La carte des thèmes dans le corpus entier



tend à rassembler les mots qui s'attachent à l'humanité (*peuple*, *langage*, *esprit*, *actes*, *science*, *poème*, *chant*, *phrases*). Mais on voit mal, malgré le relief donné au mot *morts*, ce qui distingue les mots en rouge. En réalité la matière

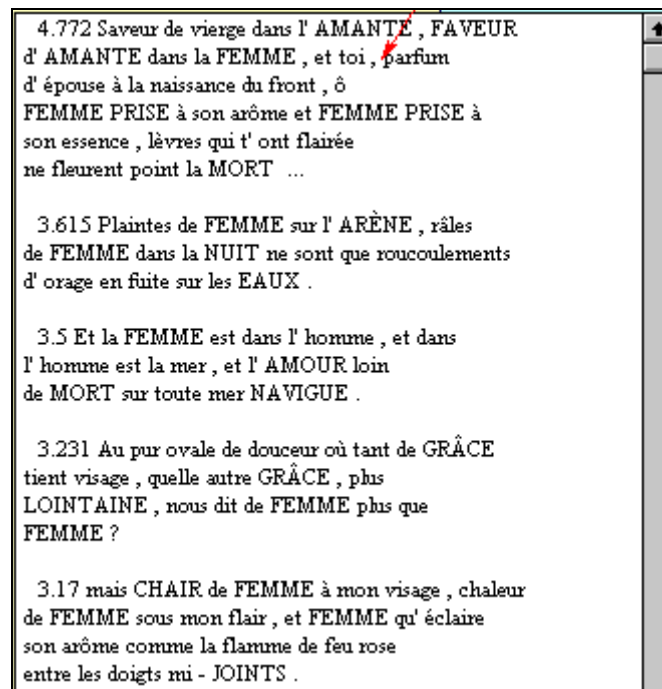
poétique est peu différenciée, le même stock lexical sert à tous les usages, comme les couleurs sur la palette du peintre.

Si l'identité des recueils n'apparaît qu'en filigrane dans la figure 16, elle peut s'affirmer au contraire quand on fait le calcul très classique des spécificités, c'est à dire des éléments lexicaux auxquels un recueil fait plus volontiers appel. Nous prendrons pour exemple le texte d'*Amers* (figure 17). La colonne de gauche consacrée aux graphies et celle de droite vouée aux lemmes s'accordent pour retenir deux thèmes principaux dans le recueil : l'amour (*amour, amant, aimer, femme, heureux, tu, te toi, moi, nous, notre*) et la mer (*mer, vagues, vaisseaux, eaux, brise*). Le thème amoureux se retrouve aussi dans la figure 18, qui met en exergue les phrases les plus caractéristiques du recueil. Ce sont celles où se trouve la plus grande concentration des termes les plus spécifiques. Elles constituent un hymne à la femme. Ces listes ordonnées de mots-clés et d'extraits typiques s'accordent le plus souvent avec l'intuition et le sentiment littéraire du lecteur.

Figure 17. Le vocabulaire spécifique dans *Amers*

Amer (forme)					Amer (lemme)				
N°	écart	corpus	texte	mot	N°	écart	corpus	texte	mot
6	27.08	530	448	mer	6	25.11	588	466	mer 2
6	10.09	152	112	toi	6	11.94	64	64	amant 2
6	9.52	175	121	tu	6	10.10	152	112	toi 5
6	8.64	44	42	es	6	9.52	175	121	tu 5
6	8.32	34	34	amante	6	9.41	271	168	ton 7
6	6.97	108	73	femme	6	6.92	93	65	amour 2
6	6.92	93	65	amour	6	6.28	157	93	te 5
6	6.10	865	386	nous	6	6.11	865	386	nous 5
6	6.05	307	158	ses	6	6.05	29	26	vague 2
6	5.93	82	55	ta	6	5.69	44	34	vaisseau 2
6	5.84	54	40	celle	6	5.34	19	18	drame 2
6	5.60	67	46	t'	6	5.15	18	17	sacré 3
6	5.55	26	23	vague	6	5.14	38	29	aimer 1
6	5.41	16	16	amants	6	5.11	73	47	ouvrir 1
6	5.34	19	18	drame	6	4.78	106	61	or 2
6	4.97	130	73	ton	6	4.55	581	254	notre 7
6	4.93	41	30	foule	6	4.54	89	52	loin 6
6	4.74	68	43	tes	6	4.49	20	17	étroit 3
6	4.54	89	52	loin	6	4.48	38	27	divin 3
6	4.37	11	11	amant	6	4.45	46	31	foule 2
6	4.29	26	20	divine	6	4.31	3284	1253	et 8
6	4.28	3284	1253	et	6	4.14	13	12	arène 2
6	4.17	52	33	te	6	4.14	10	10	scène 2
6	4.14	13	12	arène	6	4.13	78	45	moi 5
6	4.14	10	10	scène	6	4.10	187	92	mari 2
6	4.12	78	45	moi	6	4.02	33	23	voile 2
6	4.09	112	60	or	6	3.99	15	13	celle 2
6	4.01	438	192	nos	6	3.83	120	62	eaux 2
6	3.93	24	18	brise	6	3.81	938	380	son 7
6	3.90	26	19	vaisseaux	6	3.71	27	19	offrande 2
6	3.89	17	14	houle	6	3.69	16	13	heureux 3
6	3.83	120	62	eaux	6	3.63	72	40	aussi 6

Figure 18. Les phrases spécifiques dans *Amers*



Mais la garantie de la lecture fait défaut lorsqu'il s'agit de codes grammaticaux ou de structures syntaxiques, qui peuvent échapper à la conscience du récepteur et même du rédacteur et qui acceptent, comme les graphies et les lemmes, les calculs de spécificité. Ainsi dans *Amers* la catégorie du féminin est-elle privilégiée, et celle du pluriel dans *Vents*.

Quant aux structures rythmiques ou syntaxiques, leur approche est délicate, et leur récurrence n'est pas toujours sensible, même aux oreilles les plus fines. L'ordinateur qui maintient son attention sans défaillance sur de très longues distances peut reconnaître un rythme obsédant, comme celui d'un verset octosyllabique de la *Genèse*: "*Et l'Esprit planait sur les eaux*". La structure syntaxique qui sous-tend ce verset est codée sous la forme *cdnvsdn*, c'est à dire *coordination + déterminant + nom + verbe + préposition + déterminant + nom*. Elle est reconnue comme spécifique du recueil *Amers*, où se trouvent concentrés 16 des 32 exemples relevés chez Saint-John Perse et restitués dans la figure 19 (les exemples extraits de *Amers* ont le code *Am* sur la marge gauche).

Figure 19. Un rythme obsédant dans Amers

Forme	Lemme	Cout	Nb	32	CONCORDANCE
			Clic sur une ligne pour voir la page		
Am	379a				et le ciel chasse sur ses ancras . Les h
An	83a				Mais l' Étranger vit sous sa tente honoré
Ex	118a				Et la fraîcheur est dans nos murs où l'
Am	267a				et sa défense est dans mon chant Commande
Am	300a				Et l' impatience est sur les eaux du mot
Am	304a				Et la douceur est dans le chant non dans
Am	305a				Et la douceur est dans l' attente non dans
Am	320a				Et mon amour est sur les mers et ma brûlure
Am	320a				et ma brûlure est sur les mers Étroits son
Am	322a				et l' inquiétude habite sous mon sein Parf
Am	368a				Et la femme est dans l' homme et dans l'
Ch	415a				Et la roue tourne entre nos mains comme
Ve	174a				Et l' amour sanglotait sur nos couches noc
Am	325a				et l' amour gronde dans les conques comme
Ro	53a				Et l' homme marche dans les songes et s' a
Am	390a				Car ton plaisir est dans la masse et dans
An	63a				Et l' Ambassadeur mange à la table des Roi
Ex	131a				Et le salut soit sur la face des terrasses
Am	390a				mais ton délice est à la pointe du récif
Am	300a				Et la Mer lave sur la pierre nos yeux brûl
Am	276a				Et l' ombre passe d' une voile aux lisières
Am	330a				et ton haleine est dans l' exhalaison des
Am	347a				et la maison respire dans ses cloisons de

On observera avec intérêt que la plupart de ces exemples sont également des octosyllabes et que le souvenir du verset biblique est reconnaissable dans les mots même: *est sur les eaux, est sur les mers*. Ainsi quand le rythme se fait souple et abandonne la mesure répétitive et trop prévisible du vers classique, l'ordinateur peut se substituer au métronome défaillant. On imagine quel profit pourrait en tirer l'étude du verset claudélien.