



HAL
open science

Statistiques rimbaldiennes

Etienne Brunet

► **To cite this version:**

Etienne Brunet. Statistiques rimbaldiennes. S.I.@.T. 2004 : Les littératures de l'Europe unie, Sep 2004, Cesenatico, Italie. pp.88-113. hal-01362731

HAL Id: hal-01362731

<https://hal.univ-cotedazur.fr/hal-01362731>

Submitted on 9 Sep 2016

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.

Statistiques rimbaldiennes¹

Etienne Brunet

La statistique aime les grands nombres et les grands espaces. Rimbaud aimait aussi ce qui est grand : en quête des paysages d'Europe, d'Asie ou d'Afrique, il a rêvé le monde entier avant de se lancer dans l'exploration. En quête de connaissance, il a abordé la plupart des disciplines, littéraires, artistiques, techniques, en dévorant les dictionnaires et les manuels. En quête de communication, il a maîtrisé beaucoup des langues européennes dont l'anglais, l'allemand et le russe et certaines des langues africaines. En quête d'écriture, il a lu et assimilé les auteurs anciens et modernes et fait le tour de toutes les écoles littéraires, de Virgile à Baudelaire. En quête de fortune, il s'est essayé à divers gagne-pain, de l'enseignement au commerce. Il fait songer à un Bouvard qui serait fulgurant ou à un Pécuchet qui serait génial.

Mais tous ces dons accumulés ont amassé une fortune bien maigre et produit une œuvre plus mince encore, dont la postérité a assemblé religieusement les débris épars, encore fragmentaires, et dont la partie proprement littéraire atteint à peine 40 000 occurrences, soit, en surface imprimée, l'équivalent d'un petit roman, ce que Balzac ou Hugo auraient écrit en une ou deux semaines. Certes d'un bord à l'autre de cette œuvre, entre les poèmes du collégien et les dernières illuminations, la distance est considérable, mais elle est parcourue par un éclair, dans un désert où la matière manque. Or la statistique a besoin de matière pesante, de grandes masses lexicales à remuer, à filtrer et à comparer. Peu habile à

1. Actes des Séminaires d'Informatique Appliquée à l'Analyse Textuelle, SI@T, Université de Bologne, 2003-2004. Seule la mise en page a été revue et corrigée.

« peser des œufs de mouche dans des balances de toiles d'araignée ²», elle propose une machinerie puissante qui déblaise le terrain dans les grands chantiers, là où les ressources de la mémoire humaine atteignent leur limite. Même à un Balzacien reconnu, il est impossible de situer et de retrouver, de mémoire, chacun des personnages, et, à plus forte raison, chacun des termes ou expressions figurant dans la *Comédie humaine*. Balzac lui-même se perdait quelque peu dans le labyrinthe de sa création. Mais quand il s'agit de Rimbaud, certains fidèles de l'auteur sacré pourraient réciter sans faille la suite de ses poèmes et détailler les gloses que chaque mot du texte a pu susciter. Les outils documentaires sont ici de peu d'intérêt puisque les données tiennent dans un même cerveau. Et la statistique a peu de prise quand les effectifs sont faibles et son rendement est alors amoindri par la difficulté à réduire les écarts aléatoires.

Au reste l'emploi des outils documentaires et statistiques ne va pas sans une certaine naïveté qui donne sa foi aux mots imprimés, dans leur innocence première. Or chez Rimbaud les mots sont souvent pipés. Ce sont des leurres, des prête-noms, et la réalité qu'ils désignent et qu'ils cachent se dérobe parfois aux plus savantes exégèses. Quand l'ésotérisme multiplie les pièges, comment s'assurer de la constance sémantique des termes ? Bien des choses sont bleues chez Rimbaud, jusqu'à la lettre *o* (ou le son *o* ?). Mais s'agit-il simplement de la couleur physique, ou d'un état dans le processus alchimique, ou bien de quelque correspondance d'un autre ordre ? Il est alors difficile de réunir et de mélanger tous les bleus, ton sur ton, en ignorant le contexte dont ils reçoivent un reflet changeant. Ce grief, qu'on oppose depuis toujours aux dénombrements réducteurs, est plus pertinent encore dans le cas de Rimbaud.

Et pourtant c'est sur Rimbaud que la statistique lexicale a essayé ses premières armes, il y a près de cinquante ans, dans un chapitre de *Problèmes et méthodes de la statistique linguistique*. À vrai dire Guiraud avait déjà fourbi ses instruments de mesure en rédigeant sa thèse sur le vers de Valéry. Dans les deux cas la raison du choix est à trouver dans l'étroitesse du corpus. En ce temps-là l'ordinateur était réservé aux applications scientifiques ou militaires et c'est à peine si les cartes perforées permettaient un traitement mécanographique aux littéraires audacieux – tels le père Busa en Italie, le LASLA à Liège, Ch. Muller et B. Quemada à Besançon – qui commençaient à se livrer à ce travail de bénédictin. Mais Guiraud était de la race des impatientes – celle de

2. L'expression est de Voltaire et visait Marivaux.

Rimbaud – et il lui fallait essayer la méthode sans attendre, et donc sans machine. Il a donc fait les comptages à la main, en réduisant autant que possible la taille du corpus et en utilisant les bonnes volontés de son entourage. Chaque fois qu’il pouvait, il utilisait des raccourcis, en comptant les syllabes plutôt que les lettres, en écartant les mots grammaticaux des relevés et des calculs et en livrant ses conclusions avant même que les comptages soient terminés, ce qu’il reconnaît avec franchise : « [...] les chiffres donnés sont approximatifs, l’index complet n’ayant pas été complètement établi »³.

Les données n’ont donc pas la précision attendue. Les normes non plus. Pour apprécier la richesse lexicale de chacun des recueils, Guiraud se fonde sur une table expérimentale et fragile, qui enregistre la correspondance observée au cours des sondages entre le nombre de mots-occurrences (N) et celui des mots-vocables (V)⁴. Quant aux conclusions, elles mêlent le vrai à l’erreur et le contestable au vraisemblable.

Le premier problème abordé – beaucoup ont suivi cette voie – est celui de la richesse lexicale. Après avoir dressé la courbe des œuvres de Rimbaud sur une échelle bilogarithmique (N en abscisse et V en ordonnée), Guiraud observe : « Ma première pensée en établissant ce graphique était qu’il pourrait révéler un enrichissement progressif du vocabulaire, permettant de dater les *Illuminations*. Ce n’est pas le cas. Les différents moments de l’œuvre font apparaître une grande stabilité [...]. Dans ce schéma toutefois les *Illuminations* occupent une place à part [...]. Relevons en attendant la grande richesse quantitative du vocabulaire total ; il faut donc, puisque ce vocabulaire est d’étendue normale au niveau de chaque œuvre séparée, qu’il se renouvelle anormalement d’une œuvre à l’autre⁵. »

Personne n’avait alors songé à appliquer la loi binomiale au calcul de la richesse lexicale, ce que Charles Muller et Etienne Evrard allaient proposer quelques années plus tard. Cette méthode rigoureuse est maintenant adoptée par la communauté scientifique et permet d’obtenir la courbe évolutive de la figure 1.

On voit que la plupart des textes dépassent la limite (en pointillés) qui marque la zone centrale où le hasard pourrait être invoqué. Les

3. P. Guiraud, *Problèmes et méthodes de la statistique linguistique*, Paris, PUF, 1960, p. 128, note.

4. Cette table apparaît page 90 du même ouvrage.

5. *Problèmes et méthodes ...*, *op. cit.*, p. 128.

poésies de 1872 et les *Illuminations* (dans la moitié supérieure) se signalent par la richesse de leur vocabulaire et repoussent les autres textes, et particulièrement la correspondance, dans la zone des déficits.

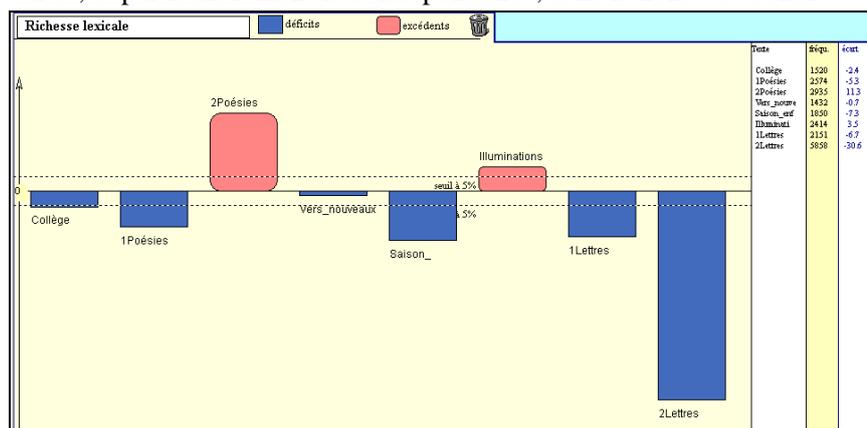


Figure 1. La richesse lexicale mesurée dans l'œuvre de Rimbaud

Négligeons ces lettres d'Europe et d'Afrique qui obéissent à d'autres normes et intéressent une autre époque, pour ne considérer, comme Guiraud, que l'œuvre poétique. Il n'est pas certain que, même ainsi ajusté, notre corpus⁶ soit superposable au sien, car certaines pièces, découvertes depuis lors, manquent dans l'édition dont on se servait au temps de Guiraud. Les résultats sont pourtant grossièrement semblables et plus facilement interprétables que n'estimait Guiraud : là où il voyait la norme respectée, nous voyons des écarts non aléatoires, dont le plus surprenant est relatif aux *Illuminations*. La prose a généralement un handicap face aux vers, ce qui se vérifie ici dans le déficit de la *Saison en enfer* et, chez Baudelaire, dans les *Poèmes en prose* comparés aux *Fleurs du mal* (voir figure 2). Que les *Illuminations* surmontent ce handicap, voilà qui en dit long sur leur exceptionnelle richesse lexicale.

6. En réalité nous avons successivement exploité plusieurs corpus, constitués différemment avec les textes de Rimbaud. Afin que nos chiffres puissent être contrôlés, nous utiliserons dans la suite de cet article les données textuelles que l'université de Paris 8 offre à la communauté scientifique et que tout un chacun peut télécharger. L'équipe qui a constitué cette bibliothèque électronique donne de grandes garanties car elle compte dans ses rangs des spécialistes voués depuis longtemps au traitement informatique et statistique des textes : Gérard Verroust, Jean-Pierre Balpe, Jean Clément, Alain Lelu. Adresse du site : <http://hypermedia.univ-paris8.fr>.

Quant à apprécier l'ensemble du corpus, la loi binomiale ne le permet pas puisqu'elle s'appuie précisément sur cet ensemble pour mesurer les sous-ensembles. Force est de recourir, soit à une norme extérieure, comme l'ensemble des textes poétiques rassemblés dans FRANTEXT, soit à quelque barème fixe ou à quelque formule à peu près fiable. Pour apprécier plus finement l'analyse de Guiraud, utilisons ses propres données en leur appliquant l'indice w . Rappelons que la formule w , qui met en relation N et V et s'écrit : $w = N V^a$ (pour $a = -0.172$), évolue en raison inverse de la richesse lexicale et que l'indice le plus bas (ici celui des *Illuminations*) dénote la richesse la plus haute.

	N	V	w
Vers (1869-1870)	6000	1590	11.57
Vers (1870-1871)	6000	1630	11.44
Vers (1872)	3000	1050	11.24
<i>Saison en enfer</i>	7225	1855	11.41
<i>Illuminations</i>	8215	2239	10.93
<i>Corpus Rimbaud</i>	30000	4420	11.39

L'indice de l'ensemble (11.39) se maintient au niveau des recueils individuels. Guiraud a donc raison de souligner le renouvellement constant de l'inspiration rimbaldienne, qui ne fléchit pas au cours du temps. Le dépérissement de l'inspiration se produit en retour chez Verlaine, comme le montre la courbe de son évolution (figure 2).

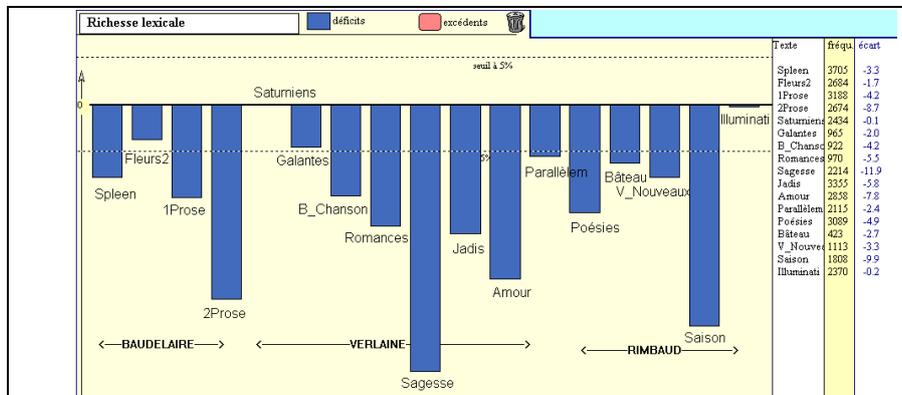


Figure 2. La richesse lexicale chez Baudelaire, Verlaine et Rimbaud (calcul sur les lemmes)

Il peut paraître cruel et déloyal de s'appesantir sur une étude vieille de cinquante ans et dépourvue des ressources que la technique offre maintenant à la recherche. Notre intention est pourtant de souligner l'intuition d'un esprit original qui reste l'un des pionniers de la discipline. Parmi les multiples indices qu'on peut appliquer à la comparaison des textes, Guiraud dans son étude sur Rimbaud en propose deux : l'un

repose sur la longueur moyenne du mot, qui va croissant chez Rimbaud, l'autre est syntaxique et mesure le rapport entre deux constructions : celle où l'adjectif accompagne le substantif et celle qui a recours à la détermination (substantif + préposition + substantif), le rapport tendant à se faire plus favorable à la seconde construction chez Rimbaud. Afin d'obtenir une assise plus large, nous reprenons notre corpus élargi, où Rimbaud voisine avec Baudelaire et Verlaine. Voici les mesures :

	Longueur rapportcoefficients de variation			
	en nombre de lettres	adjectiv./détermination	longueur	rapport syntaxique
BAUDELAIRE				
Spleen et idéal	5.09	8.06	0.44	0.48
Fleurs du mal 2	5.13	5.43	0.37	0.84
Poèmes en prose 1	5.25	4.98	0.40	0.74
Poèmes en prose 2	5.11	4.72	0.55	0.74
VERLAINE				
Poèmes Saturniens	5.08	4.32	0.43	0.94
Fêtes Galantes	5.05	4.51	0.70	0.85
Bonne Chanson	4.92	6.44	0.27	0.68
Romances s. p.	4.94	5.57	0.41	1.08
Sagesse	4.83	5.16	0.43	0.64
Jadis et naguère	4.99	4.29	0.48	0.77
Amour	4.90	4.72	0.39	0.73
Parallèlement	4.97	4.25	0.46	1.06
RIMBAUD				
Poésies	5.08	3.95	0.52	0.96
Bateau_ivre	5.26	2.41	0.65	0.24
Vers Nouveaux	4.87	2.83	0.41	0.91
Saison en enfer	4.89	2.90	0.37	0.81
Illuminations	5.22	2.31	0.46	1.07

Tableau 1. Indice lexical et indice syntaxique

Comme le remarque Guiraud le mot a tendance à s'allonger dans les *Illuminations* (alors qu'il se raccourcit de plus en plus chez Verlaine) mais la dispersion des valeurs n'y est pas plus forte qu'ailleurs et l'on ne peut pas en tirer argument pour prouver l'hétérogénéité chronologique de ce recueil. Mais il est vrai que le calcul de Guiraud portait sur le nombre de syllabes et seulement sur les mots forts (mots grammaticaux exclus). L'analyse factorielle de la figure 3 illustre le détail de la répartition en classes de n lettres (les mots très longs étant regroupés dans la dernière classe $lg11$). Une fois de plus on constate l'irrédentisme des mots grammaticaux, qui ont un statut à part et qui, refusant de prendre parti entre les mots courts et les mots longs, s'établissent sur la ligne de partage, c'est-à-dire sur l'axe des x ($lg1$, $lg2$ et $lg3$)⁷. Les classes qui sont situées au-delà de 3 lettres, et où les mots grammaticaux deviennent rares, se répartissent selon un arc de arc régulier qui commence par Verlaine

7. L'ostracisme de Guiraud à leur endroit se justifie donc.

(mots courts de 4 ou 5 lettres), pour rejoindre la poésie versifiée de Rimbaud et de Baudelaire (mots de 7, 8 et 9 lettres) et prendre fin dans la prose poétique, qu'elle soit de Rimbaud ou de Baudelaire (mots de plus de 9 lettres).

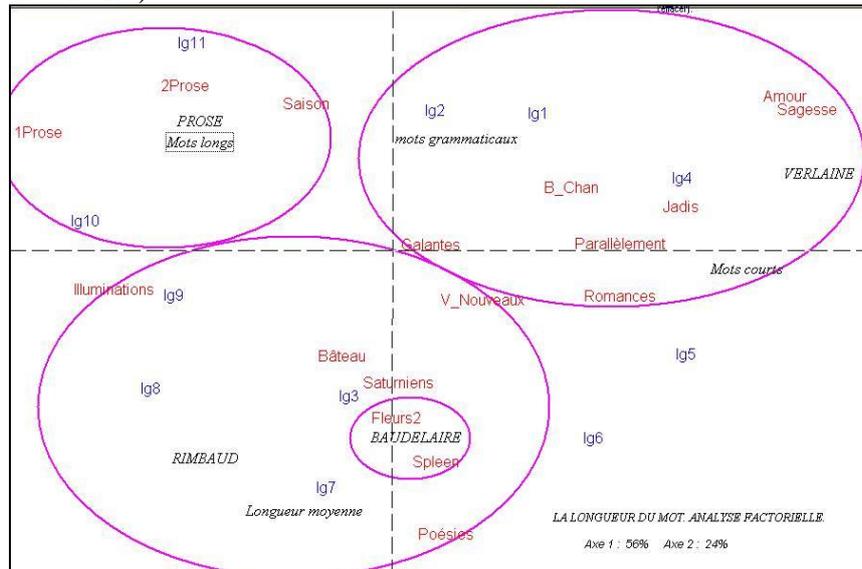


Figure 3. Analyse factorielle de la longueur des mots
(les classes lg1, lg2, lg3 regroupent les mots qui ont 1, 2, n lettres)

Le second indice résiste mieux à l'épreuve : même si le codage grammatical a été manuel dans un cas et automatique dans l'autre⁸. Non seulement l'évolution de Rimbaud lui fait préférer de plus en plus le recours au substantif plutôt qu'à l'adjectif quand il s'agit d'introduire une qualification (le rapport adjectif/substantif dans une telle situation évolue de 3.95 à 2.31), mais c'est aussi dans les *Illuminations* que ce rapport varie le plus largement, ce qui invite à étendre la fourchette chronologique où l'on peut situer la rédaction de ce recueil. Ce critère syntaxique est efficace non seulement pour dater les œuvres d'un même écrivain, mais aussi pour distinguer les écrivains les uns des autres, et opposer les genres les uns aux autres. Sa sensibilité simultanée aux trois réactifs que sont l'auteur, le genre et l'époque, en rend toutefois l'interprétation délicate, ce qui est malheureusement le cas pour bien d'autres critères.

8. Nous avons eu recours à *Cordial Analyseur* pour la lemmatisation du corpus.

Pour donner plus d'ampleur à l'intuition et à la démonstration de Guiraud, nous ouvrons le champ de l'expérimentation à la trilogie précédemment abordée. La figure 4 rend compte du couple adjectif-substantif (en réunissant les effectifs où l'adjectif est antéposé ou postposé), tandis que la figure 5 illustre la répartition du tricode substantif+préposition+substantif⁹, parmi beaucoup d'autres séquences syntaxiques. Le déclin de la paire adjectif+substantif s'observe effectivement chez Rimbaud, et aussi le progrès de la séquence substantif+préposition+substantif. Mais ces deux structures ne sont pas liées pareillement chez Verlaine, chez lequel adjectivation et détermination décroissent en même temps. Les combinaisons syntaxiques ont une complexité telle qu'un seul indice ne saurait en rendre compte¹⁰. Aussi avons-nous envisagé d'autres séquences syntaxiques, réunies dans la même analyse factorielle (figure 5).

9. Quand les procédures automatiques ou manuelles sont engagées, il est rare qu'on réduise à la fois le bruit et le silence. On peut craindre des oublis et des lacunes quand le repérage est manuel, silences que réduit l'automatisme, mais inversement une sélection automatique insuffisamment précise laisse passer beaucoup d'intrus, parmi lesquels il faut ici compter toutes les séquences où la préposition ne met pas en rapport les substantifs qui l'entourent.

10. Guiraud s'en rendait compte le premier. Et dans le même article il recommande d'« établir une liste de caractères statistiques (phonétiques, lexicaux, grammaticaux) assez nombreux » et de calculer « ensuite la corrélation des rangs pour tous les poèmes pris deux à deux ». Mais faute de moyens, il a reculé devant la tâche : « C'est évidemment, entre l'établissement des faits et les calculs, un travail gigantesque » qu'il réserve aux machines futures. « On pourrait songer à le faire avec une machine à calculer et cela devrait donner un résultat. »

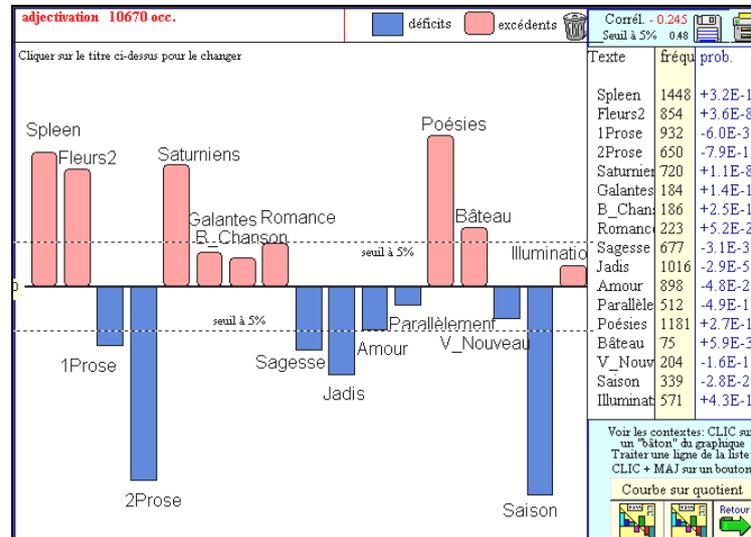


Figure 4. Les séquences adjectif+substantif et substantif+adjectif

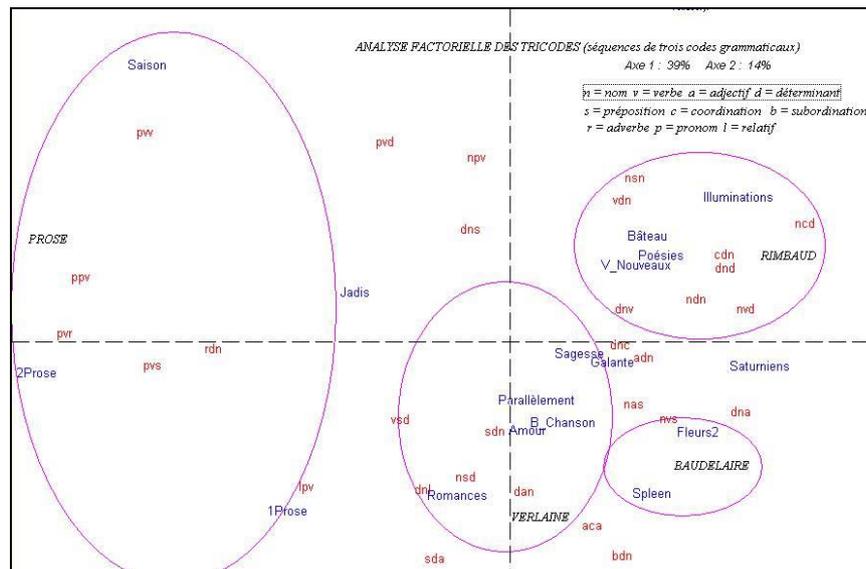


Figure 5. Analyse factorielle des trinomes (séquences de trois codes grammaticaux)

On repèrera dans l'entourage des *Illuminations* la séquence *nsn* de la détermination (substantif + préposition + substantif), en haut et à droite du graphique. L'intuition de Guiraud est juste qui voit en cela une caractéristique des *Illuminations*. Mais c'est plus largement une

caractéristique de Rimbaud, dont toutes les œuvres sont groupées en cet endroit du graphique, au voisinage d'autres structures syntaxiques qui toutes comportent au moins un substantif, associé le plus souvent à un déterminant. Les vers de Baudelaire et de Verlaine se partagent la moitié inférieure du graphique, où affluent toutes les combinaisons où entre l'adjectif. Quant à la moitié gauche, elle est monopolisée par les textes en prose, qu'ils soient de Baudelaire ou de Rimbaud. Les configurations qui apparaissent dans ces paragraphes comportent exclusivement des verbes, des adverbes et des pronoms.

Cette polarisation des parties du discours, reconnaissable dans les combinaisons syntaxiques, mérite donc une enquête directe, qu'on a réalisée dans la figure 6. Tous les mots et signes du corpus ont reçu de *Cordial* une première étiquette grammaticale qui indique sa nature¹¹. À ce stade les erreurs ne sont pas très nombreuses, l'analyse étant moins fiable quand elle se poursuit ensuite dans le maquis des homographes verbaux, où les modes, les temps et les personnes revêtent souvent des formes semblables. C'est au total 18 espèces grammaticales (y compris les signes de ponctuation répartis en deux groupes, fort et faible) dont on suit le dosage, de Baudelaire à Rimbaud. La carte obtenue, où tous les mots ou signes sont pris en compte, est superposable à la précédente. La prose poétique, à gauche, affirme sa spécificité verbale. C'est le verbe, auxiliaire ou principal, qui s'accorde avec le caractère narratif des poèmes en prose et de la *Saison en enfer*¹².

11. L'analyse du lemmatiseur ne se borne pas à cette seule catégorisation : s'il s'agit d'un verbe des précisions sont ajoutées qui identifient le mode, le temps, la personne, ou bien le genre et le nombre s'il s'agit d'une catégorie nominale. Les mots grammaticaux, variables ou non, reçoivent aussi des codes appropriés.

12. La prose n'est pas nécessairement liée au verbe. Tout dépend du genre. S'il s'agit de prose technique, le verbe se fait rare et le substantif est alors le principal support de l'information. S'il s'agit de récit, de communication orale ou de théâtre, alors le verbe reprend ses droits. Dans le corpus présentement étudié, on a affaire à la prose poétique, et le choix du verbe ou de la classe nominale est fonction de la situation du discours : si le texte est narratif, le verbe aura sa place, comme dans *Une saison en enfer*. S'il s'agit d'évocation ou de paysage, comme dans les *Illuminations*, la classe nominale aura la préférence.

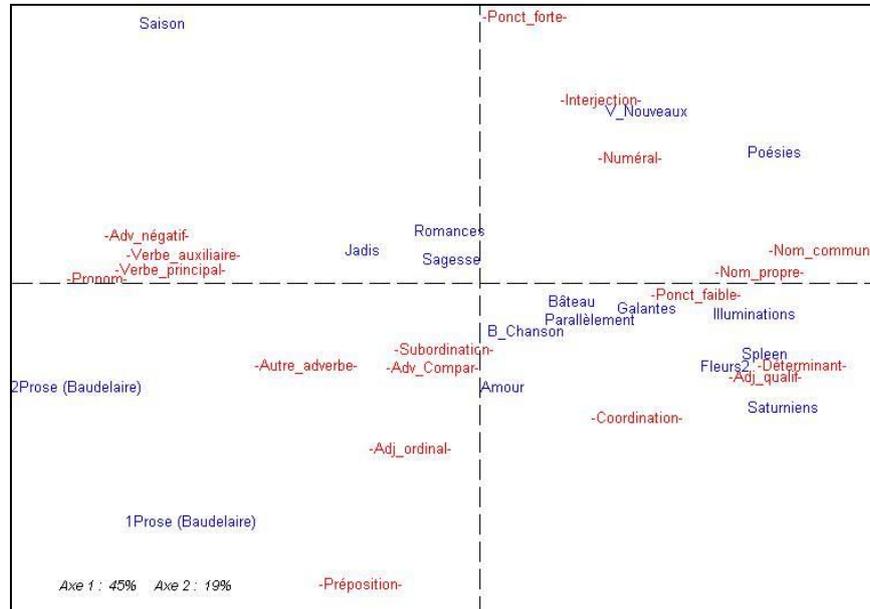


Figure 6. Analyse factorielle des parties du discours

Le verbe attire auprès de lui ses familiers, principalement les pronoms et les adverbes. Le clan opposé est celui de la catégorie nominale, qui occupe la partie droite, les noms, propres ou communs, se situant à l'extrême, avec les déterminants. L'adjectif qualificatif, comme d'habitude, fait alliance avec le substantif, et s'accorde à la tendance descriptive qui anime souvent les poèmes.

Mais si les trois poètes se situent dans le camp nominal, ils maintiennent leurs distances avec le second facteur qui oppose le haut, où s'installe Rimbaud, en accaparant l'interjection, les numéraux et les ponctuations fortes, et le bas où Baudelaire, qui voisine avec Verlaine, accueille favorablement l'adjectif qualificatif. Le discours de Rimbaud paraît ainsi plus heurté, plus violent, celui de Baudelaire plus fluide et plus suggestif. Quant à Verlaine, si ses premiers recueils *Poèmes Saturniens* et *Fêtes galantes* ne se démarquent pas des *Fleurs du mal*, son indépendance tend à s'affirmer ensuite.

La segmentation de l'oeuvre de Rimbaud ne s'impose pas d'elle-même. Mis à part la *Saison en enfer*, seul livre publié par Rimbaud, les éditeurs ont pris la liberté de classer les pièces selon la date présumée de la rédaction et de les répartir dans un ou plusieurs recueils. Guiraud en distingue cinq. D'autres fondent en un seul volume toutes les pièces en

vers antérieures à 1872, ce que nous avons fait aussi dans les pages qui précèdent, en détachant cependant le poème le plus connu, le *bateau ivre*, pour isoler l'objet du présent colloque. Il n'y a guère d'incertitude pour ce poème qui semble avoir été écrit pour être montré à Paris en septembre 1871 et qui y fut effectivement admiré. Mais faisons comme si ce poème avait été trouvé sans date ni nom d'auteur, ce qui est d'ailleurs le sort infortuné de quelques écrits de Rimbaud. Pourrait-on l'attribuer à Rimbaud en s'appuyant sur l'analyse statistique et sur une idée émise par Guiraud dans l'article étudié : « On pourrait établir un tableau de corrélations lexicales entre les différentes œuvres pour voir les mots qu'elles ont en commun et ceux qu'elles ont en propre. Mais c'est un travail énorme.¹³ »

C'est ce que réalise la figure 7, sans grand mérite. Car ce qui paraissait énorme en 1960 est un jeu d'enfant pour les machines actuelles. Le tableau des corrélations dont rêvait Guiraud prend la forme d'un triangle où chaque texte est pourvu d'une distance qui le sépare de chacun des autres. Ce tableau, analogue à celui que les atlas proposent pour les distances de ville à ville, est calculé sous la forme imaginée par Guiraud et connue maintenant sous le nom de Jaccard. La formule employée par notre logiciel HYPERBASE – il y a d'autres variantes – est établie comme suit :

$$d = ((a-ab)/a) + ((b-ab)/b),$$

où ab désigne la partie commune aux vocabulaires a et b des deux textes comparés ($a-ab$ et $b-ab$ recouvrant les parties privatives). Cette méthode de calcul, confirmée par d'autres approches où la fréquence entre en ligne de compte, invite à rattacher le *Bateau ivre* au recueil des *Poésies* de Rimbaud. Les autres textes du même auteur sont beaucoup plus éloignés, parfois même plus lointains que certains recueils de Baudelaire (*Spleen et idéal*) ou de Verlaine (*Poèmes Saturniens*). Il n'y a donc pas lieu de faire de l'aventure du *Bateau ivre* l'anticipation des ruptures majeures où l'on verra bientôt Rimbaud rompre toutes les amarres.

13. *Problèmes et méthodes...*, op. cit., p. 129.

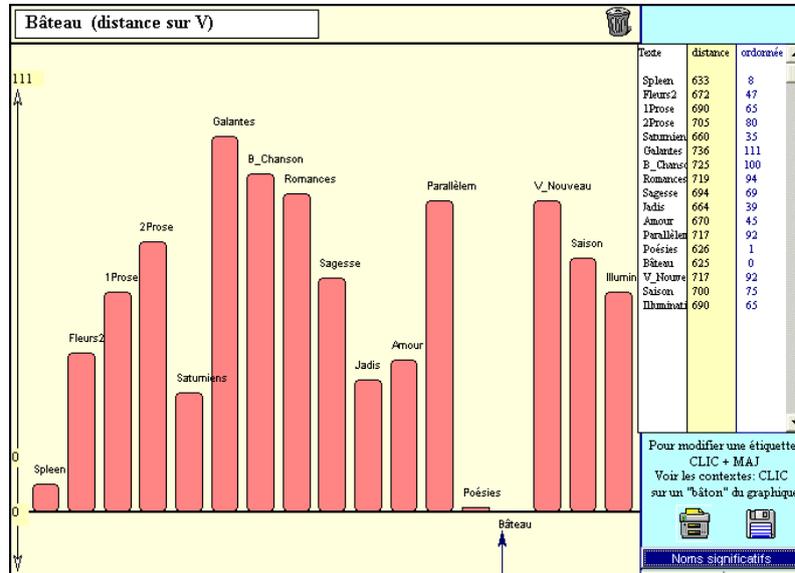


Figure 7. Courbe établissant la distance du *Bateau ivre* aux autres textes

Cependant la courbe seule qui isole le *Bateau ivre* ne dit rien du jeu d'alliances ou d'oppositions qui gouverne l'ensemble des textes comparés. À partir du tableau des distances, il est possible de dessiner la carte synthétique, que ce soit pour les villes ou pour les textes. La procédure la plus efficace est celle de l'analyse arborée qui répartit au mieux les textes dans l'espace, compte tenu des distances observées deux à deux.

La carte obtenue (figure 8) distingue nettement les trois écrivains et la typologie des textes dans l'œuvre de chacun. Verlaine se détache dans la partie basse en s'éloignant de plus en plus. Son parcours suit la chronologie : si les premiers recueils (*Poèmes saturniens*, *Fêtes galantes* et *Romances sans paroles*) sont encore proches de Baudelaire et du premier Rimbaud, l'inspiration change avec la *Bonne Chanson*, qui annonce les thèmes et la manière de *Sagesse* et de *Amour*. Cependant Verlaine est sujet aux regrets, aux rechutes et aux repentirs et certains recueils tardifs (*Jadis et naguère*, *Parallèlement*) témoignent de cette inconstance. Chez Baudelaire ce n'est pas la chronologie mais le genre qui explique l'apparition de deux branches séparées, l'une conduisant aux vers, l'autre à la prose.

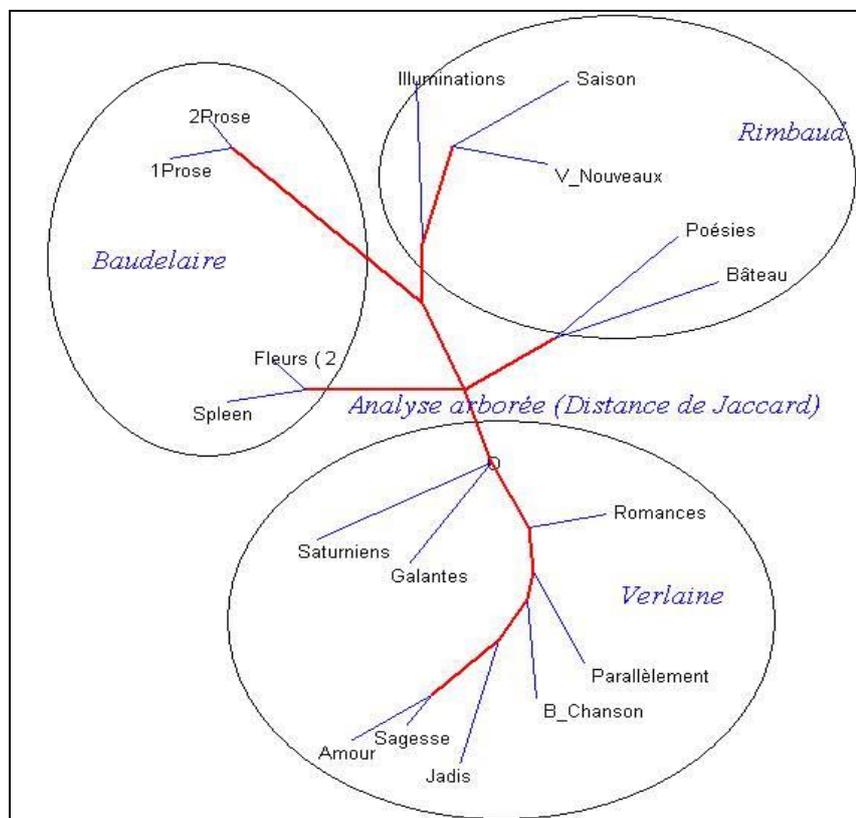


Figure 8. Analyse arborée de la distance lexicale

Il n'en est pas ainsi chez Rimbaud, où cohabitent pourtant vers et prose. La première branche¹⁴ est certes homogène, et ne contient que des pièces en vers, celles du premier recueil. Mais la seconde inclut des œuvres en vers (*Vers Nouveaux*) et en prose (*Saison en enfer* et *Illuminations*), ce qui laisse penser que la rupture est dans le temps et non dans le genre. Au reste certaines pièces en vers subsistent dans la prose de la *Saison en enfer*, et le vers libre apparaît fugitivement dans les *Illuminations*¹⁵. On a tout lieu de penser que la production de Rimbaud à partir de 1872 mêle indistinctement vers et prose et que l'on passe

14. Le premier recueil de Rimbaud est le plus proche des *Fleurs du mal*. L'influence du grand devancier s'exerce sur les débuts de Rimbaud comme sur ceux de Verlaine.

15. Un des textes les plus obscurs des *Illuminations* est intitulé *Sonnet*, sans qu'on sache trop pourquoi.

insensiblement de l'un à l'autre, sans que le vocabulaire, la syntaxe et les thèmes en subissent les effets.

L'analyse qui précède porte sur les formes, où s'entremêlent les faits qui tiennent au contenu thématique et ceux qui s'attachent au style et à la syntaxe. Comme on a procédé à la lemmatisation du corpus, les lemmes, en regroupant les formes, perdent la plupart de leurs caractéristiques grammaticales et permettent d'isoler le contenu sémantique. La figure 9 corrobore les conclusions que suggérait la figure 8, bien qu'elle mette en jeu un autre objet (les lemmes plutôt que les formes), une autre mesure (la fréquence plutôt que la présence-absence) et une autre méthode (l'analyse factorielle plutôt que arborée). Même dichotomie entre la prose et les vers chez Baudelaire. Même distinction radicale entre ce qui revient à Verlaine et ce qui appartient à Rimbaud. Même influence enfin des *Fleurs du mal* sur les débuts de Verlaine et de Rimbaud.

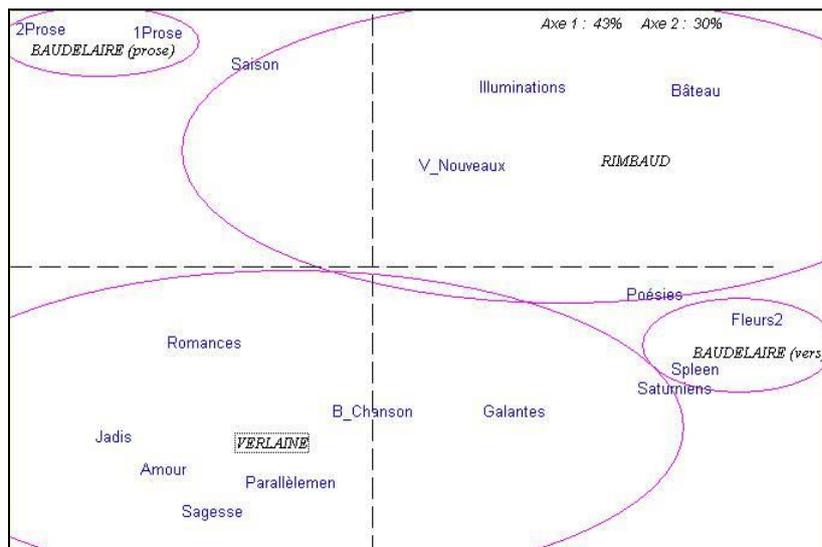


Figure 9. Analyse factorielle de la distance lexicale (fréquence des lemmes)

Quant à la syntaxe, elle peut être isolée pareillement grâce au codage grammatical dont tous les mots du corpus ont été pourvus. Comme ce codage est assez approfondi – sans être toujours fiable – les variétés sont nombreuses. On a compté 1798 variétés différentes de ces coques vides, dont le sens a été évacué et dont le dosage dans les textes peut être mesuré. Et le calcul appliqué précédemment aux lemmes est repris dans la figure 10, les étiquettes se substituant aux mots. On rejoint ainsi l'analyse pratiquée sur les parties du discours (figure 6). Mais on va plus

loin ici dans le détail, puisque les marques du nombre, du genre, du temps, du mode, de la personne, etc. sont prises en compte.

Cette fois les différences stylistiques qui opposent la syntaxe des vers à celle de la prose sont très sensibles dans le cas de Baudelaire. Verlaine dont l'œuvre tout entière est versifiée ne se distingue pas des *Fleurs du mal* et si sa métrique innove, sa syntaxe reste classique, sauf, semble-t-il, dans le recueil *Romances sans paroles* qui s'écarte du côté de Rimbaud. Car Rimbaud se tient à distance sur le côté gauche de la figure en manifestant son originalité et son indifférence aux canons du genre : prose et vers se côtoient chez lui sans s'opposer. C'est à peine si l'on sent un léger écart dans la *Saison en enfer*, dont la position tend à de rapprocher des *Poèmes en prose* de Baudelaire.

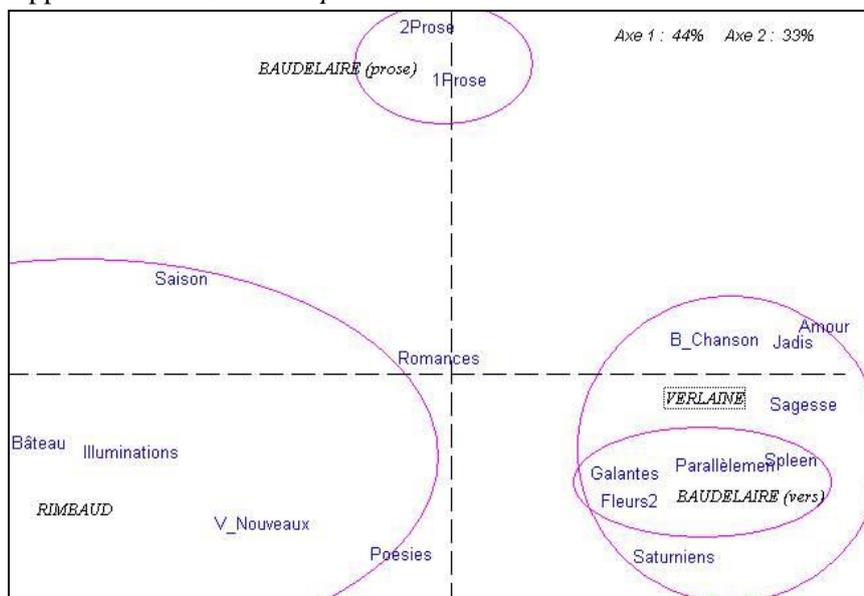


Figure 10. Analyse factorielle de la distance grammaticale

Faut-il dès lors remettre en cause cette parenté que les critiques ont soulignée et que les intéressés eux-mêmes ont revendiquée¹⁶ ? Pour en décider, il faut élargir le champ et considérer l'ensemble des écrivains présents dans FRANTEXT. Si large que soit cette base, il convient

16. Si Verlaine et Rimbaud voient en Baudelaire le grand aîné, celui-ci en retour disparaît trop tôt pour connaître ses cadets. Il meurt en 1867. Les *Poèmes saturniens* venaient à peine de paraître et Rimbaud fréquentait encore le collège.

cependant de ne retenir que les écrivains dont l'œuvre est suffisamment représentée. Rimbaud, Verlaine et Baudelaire sont de ceux-là, et quelque 67 autres, d'Honoré d'Urfé à Julien Gracq¹⁷. Le calcul de distance trouve à s'employer dans cette nouvelle base grosse de 55 millions d'occurrences et de 70 profils, dont on extrait celui de Rimbaud représenté dans la figure 11. L'auteur du *Bateau ivre* est rapproché de tous les autres écrivains et la distance qui le sépare de chacun est matérialisée par la longueur du segment. Les plus courts désignent des poètes comme Verlaine ou Leconte de Lisle. Baudelaire n'est pas loin mais son œuvre en prose, surtout la prose des *Salons*, tend à l'éloigner de Rimbaud. Il en est ainsi de Nerval et de Mallarmé. Néanmoins le trio Baudelaire-Verlaine-Rimbaud est reconstitué au bas du graphique. Il l'est aussi dans la figure 12 qui fait la synthèse de tous les profils, et dresse la carte littéraire des quatre derniers siècles.

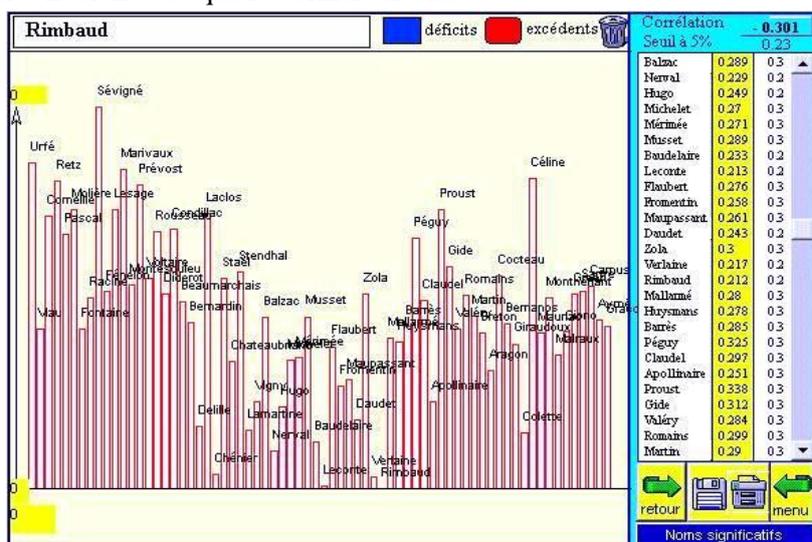


Figure 11. Distance de Rimbaud aux autres écrivains français

17. On n'a pas cru incorporer les auteurs, même importants, du XVI^e siècle, parce que l'inconstance de l'orthographe aurait faussé les comparaisons.

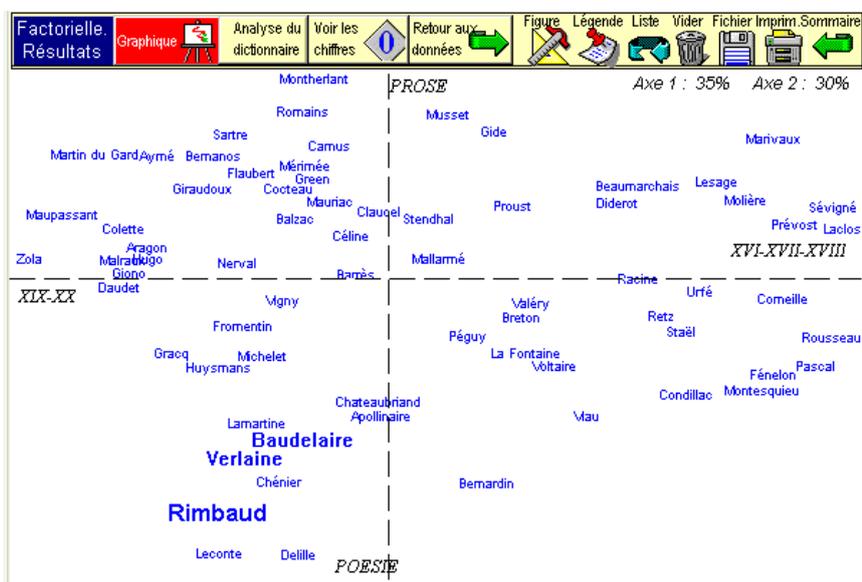


Figure 12. Analyse factorielle de la distance lexicale appliquée à 70 écrivains

La même base qui nous révèle avec quels écrivains Rimbaud fait alliance nous indique encore avec quels mots il aime à s'exprimer. Le calcul très classique des spécificités met en relief certaines caractéristiques qui appartiennent certes au lyrisme mais qui acquièrent chez Rimbaud une violence neuve. C'est le cas du point d'exclamation et de l'interjection *ô* qui occupent les deux premières places de la liste (tableau 2). On peut trouver que la thématique des fleurs (*lys, fleurs, fleurie, parfums*) est traditionnelle chez les poètes, et plus généralement celle du ciel et de la nature (*cieux, saisons, lunes, ciel, azur*). Mais ce spectacle habituel à la poésie, Rimbaud le voit avec des verres correcteurs, sensibles à la pluralité (*des, aux, les*), au mouvement (*danse, dansent, chante, élans, frissons*), et à la couleur (*noirs, bleu, bleus, verts, or, violettes, blanc*).

43.56	302065	865	!	36.70	13549	125	ô	26.44	476172	838	des
21.27	3416	36	noirs	19.88	749566	1009	les	18.76	616	13	lys
17.51	3612	31	bleu	17.35	88829	204	aux	17.08	1091	16	bleus
17.03	222	7	splendides	16.42	238	7	dansent	15.99	2427	23	cieux
15.84	969	14	verts	15.02	561	10	saisons	14.95	1552	17	chante
14.85	7138	39	fleurs	14.75	216	6	lunes	14.51	156	5	squelettes
14.44	995	13	parfums	14.40	305	7	élans	14.18	163	5	fleurie
13.86	14708	56	or	13.81	1590	16	danse	13.79	428	8	violettes
13.37	694	10	lèvre	12.93	20084	64	ciel	12.73	200	5	viendras
12.35	212	5	naïfs	12.34	5609	29	blanc	12.18	1341	13	azur

Tableau 2. Les mots spécifiques de Rimbaud parmi 70 écrivains

Si l'on élargit encore la toile de fond, en y englobant tout FRANTEXT, soit une base quadruplée de 200 millions de mots, la liste reste stable. Aucun élément du tableau 2 ne manque à l'appel, et la hiérarchie est consolidée. On observe seulement quelques additions intercalées : *bave, boissons, enfer, soif, châteaux, seins, soleil, astres*. Mais il est aussi légitime de rétrécir le champ d'observation et d'exclure les prosateurs, avec lesquels Rimbaud ne partage rien, ou les auteurs trop anciens avec qui Rimbaud n'a rien en commun. Et pour comparer ce qui est comparable, il peut être pertinent de s'en tenir à l'époque de Rimbaud et au genre poétique qu'il a cultivé. Nous reprendrons pour cela notre corpus initial, centré autour de Baudelaire, Verlaine et Rimbaud.

L'échelle étant réduite, l'objet de la focalisation peut se limiter au texte du *Bateau ivre*, si court soit-il¹⁸ (figure 13). Le thème de la navigation, imposé par le titre, est évidemment au premier plan, ballotté entre l'air (*ciel, oiseau, aube, soleil*) et l'eau (*bateau, flot, poisson, mer, île, baigner, eau, fleuve*). Mais on retrouve aussi dans ce court poème des caractéristiques qui sont celles de Rimbaud : le mouvement (*descendre, monter, courir*), la couleur (*bleu, violet, vert, jaune*) et le pluriel (*des, aux, les*), très reconnaissable dans la liste de gauche, consacrée aux formes, où quasiment tous les substantifs sont au pluriel : *cieux, flots fleuves, enfants, mers, oiseaux, nuits, fleurs*¹⁹. Les critiques ont ressassé à l'envi le caractère symbolique de ce naufrage volontaire, qui annonce

18. En de tels cas, les effectifs devenant trop petits, il est prudent d'utiliser le calcul hypergéométrique, plutôt que la loi normale. C'est cette méthode de calcul qui est employée ici, même si la liste semble ordonnée selon l'écart réduit restitué en deuxième colonne de la figure 13. En réalité cet écart réduit est le fruit d'une transformation de la probabilité issue du calcul hypergéométrique.

19. Seules exceptions : l'*Europe*, qui est unique, le *bateau* qui dit *je*, et l'*eau* qui est peu nombrable. Dans le corpus le rapport pluriel/singulier est égal à 24079/58238, soit 0,41. Il est nettement plus favorable au pluriel chez Rimbaud 7373/11477 = 0,64, et plus encore dans le *Bateau ivre* 307/205 = 1,49. Or on a constaté généralement que ce rapport était plus élevé dans la poésie, sensible à la « multiple splendeur » du monde et peut-être aussi à l'amplification propre au pluriel. De ce point de vue il y a une surenchère chez Rimbaud. Mais inversement Rimbaud s'applique à un décalage un peu provoquant, où le ton poétique est bousculé et parfois violenté par l'introduction de mots que la poésie n'a jamais connus. Passe encore pour les mots rares comme *dérades, bonaces, ultramarin* ou *flache*, qui apportent un dépaysement analogue à celui que recherche Hugo en multipliant les noms propres inconnus. Rien à objecter non plus à ces mots nobles comme *fientes, rut, voire vomissure*, dont la noblesse recouvre des réalités qui le sont moins. Mais les poètes, même ceux qui prétendent comme Hugo avoir mis un bonnet rouge au dictionnaire, n'ont pas coutume d'évoquer la *morve*, les *vacheries* et les *tohu-bohus cataractants*. Toutefois cette insolence est mesurée, et presque fondue dans le vocabulaire traditionnel de la poésie, dont rend compte la liste exhaustive établie dans le tableau 3.

celui de Rimbaud lui-même. Il est piquant de constater que les chiffres leur donnent raison : le poème le plus connu de Rimbaud, pour ses contemporains comme pour la postérité, est aussi celui qui résume au mieux les traits les plus saillants de son écriture, au point qu'il paraît non seulement le symbole de son destin, mais le concentré et la quintessence de son oeuvre.

N°	écart	corpus	texte	mot	N°	écart	corpus	texte	mot
14	6.2	2442	40	des	14	5.7	1301	26	de
14	4.8	5	3	bateau	14	4.5	7	3	bateau
14	4.3	560	13	aux	14	3.7	54	4	descendre
14	4.1	528	12	j'	14	3.7	52	4	flot
14	4.1	303	9	ai	14	3.6	5	2	poisson
14	4.1	3015	34	les	14	3.6	2029	24	je
14	3.9	42	4	vu	14	3.5	1737	21	avoir
14	3.7	4	2	baigné	14	3.4	7	2	éveil
14	3.6	5	2	taché	14	3.3	9	2	tacher
14	3.5	65	4	cieux	14	3.3	9	2	europe
14	3.4	32	3	flots	14	3.3	140	5	mer
14	3.3	9	2	europe	14	3.2	1512	18	au
14	3.0	175	5	ont	14	3.0	15	2	île
14	3.0	14	2	fleuves	14	3.0	105	4	bleu
14	2.9	18	2	béni	14	2.9	16	2	violet
14	2.8	20	2	verte	14	2.9	16	2	baigner
14	2.7	73	3	enfants	14	2.9	116	4	eau
14	2.7	22	2	ners	14	2.8	62	3	oiseau
14	2.7	22	2	ivre	14	2.8	300	6	ciel
14	2.5	91	3	eau	14	2.8	20	2	regretter
14	2.5	357	6	moi	14	2.8	20	2	frêle
14	2.5	31	2	oiseaux	14	2.6	85	3	vert
14	2.5	269	5	nes	14	2.6	29	2	peau
14	2.3	41	2	nuits	14	2.6	28	2	aube
14	2.3	115	3	ner	14	2.6	26	2	ivre
14	2.3	114	3	fleurs	14	2.5	90	3	monter
14	2.2	678	8	plus	14	2.5	356	6	moi
14	2.1	63	2	bleu	14	2.5	33	2	courir
14	2.1	59	2	parfois	14	2.5	32	2	jaune
14	2.1	246	4	ô	14	2.4	39	2	fleuve
					14	2.4	191	4	enfant
					14	2.4	186	4	soleil

Figure 13. Spécificités du *Bateau ivre* (dans le corpus Baudelaire-Verlaine-Rimbaud) (à gauche les formes, à droite les lemmes)

La couleur est si présente chez Rimbaud qu'elle peut en devenir presque lassante. Une lecture superficielle peut en effet laisser l'impression de pauvreté d'inspiration ou d'expression, comme ces cahiers de coloriage que gribouillent les enfants. La surcharge pourrait faire penser à ces pages lourdement descriptives où Zola accumule les couleurs et les adjectifs, pour peindre les Halles ou le Paradou. Il y a loin pourtant entre l'application besogneuse et zolienne qui installe pesamment son chevalet pour reproduire une « nature morte » et le trait rapide et suggestif qui évoque une vision rimbaldienne, comme celle qui a pour titre *Fleurs* dans les *Illuminations* et où les pierres précieuses viennent enrichir la palette²⁰:

20. Il existe une sorte de préciosité littéraire qui fait parler les pierres précieuses. Elle est sensible dans le recueil de Théophile Gautier *Émaux et camées* et s'illustre plus encore chez Mallarmé. Le lapis-lazuli originel brille encore dans l'azur de Mallarmé.

!	20	choient	1	exiler	1	lâcher	1	on	1	rythme	1
,	89	cible	1	exquis	1	lactescent	1	or subst.	2	sanglot	1
.	11	ciel	6	faire	2	laisser	1	or coor.	1	sans	4
:	4	cinquante	1	falot	1	lame	1	orgueil	1	sapin	1
;	4	circul.	1	femme	1	langueur	1	où	6	savoir	3
?	1	clabaudeur	1	fermentent	1	lasser	1	ou	1	sentant	1
...	1	clapotement	1	fermenter	1	laver	1	ouragan	1	sentir	1
1871	1	clouer	1	fiente	1	le pron.	2	ouvrir	1	serpent	1
à	11	colombe	1	figement	1	le art.	66	panthère	1	sève	1
accroupi	1	comme	4	filer	1	léger	1	papillon	1	si	1
âtre	1	confiture	1	finir	1	lent	1	par	3	sidéral	1
acteur	1	coque	1	flache	1	lenteur	1	parapet	1	sillage	1
aïler	1	coton	2	flamand	1	leur	2	pareil	2	soir	1
ainsi	2	couleur	1	flamme	1	léviathan	1	parfois	2	soleil	4
alcool	1	coup	2	fleur	3	libre	1	parfum	1	son	1
aller	1	courant	1	fleuve	2	lichen	1	pas	2	songer	1
amer	2	courir	2	florides	1	lien	1	peau	2	sourd	1
amour	2	crépuscule	1	flot	4	lieue	1	pénétrer	1	sous	4
ancien	1	crever	1	flottaison	1	loin	1	péninsule	1	subir	1
anglais	1	criard	1	fond	2	lointain	1	pensif	1	suivre	1
anse	1	croire	1	forcer	1	long	1	perdre	1	sûr	1
antique	1	crouler	1	fort	1	lors	1	peuple	1	sur	2
appeler	1	dans	4	fou	1	lorsque	1	phosphore	1	tache	1
arbre	1	danser	1	frêle	2	lumineux	1	pied	1	tacher	2
arc	1	de	69	frisson	1	lune	1	planche	1	tapage	1
archipel	1	délirant	1	froid	1	lunule	1	plein	1	te	1
ardent	1	délire	1	fumer	1	lyre	1	plein	1	teindre	1
argent	1	démarrer	1	furieux	1	maelstrom	1	pleurer	1	tempête	1
assaut	1	dérader	1	futur	1	mai	1	plus	8	tendu	1
astre	1	dés	1	géant	1	mais	1	poème	1	tohu	1
atroce	1	descendre	4	geindre	1	marais	1	poète	1	tordre	1
au	18	désirer	1	genou	1	marée	1	poisson	2	torpeur	1
aube	2	dévorant	1	glacier	1	maries	1	pôle	1	tout adj.	3
autre	1	dévorer	1	glauque	1	maritime	1	pomme	1	tout adv.	2
avec	3	dispersant	1	golfe	1	martyr	1	ponton	1	travers	1
avoir	21	dix	1	gonfler	1	me	6	porter	1	traverser	1
azur	2	dont	3	gouffre	1	mêlant	1	porteur	2	trembler	1
ba	1	dorade	1	gouvernail	1	mer	5	poteau	1	très	1
baigner	2	dormir	2	grappin	1	milieu	1	pour	1	triomphant	1
baiser	1	doux	2	guider	1	million	1	pourrir	1	trique	1
ballotter	1	drame	1	haleur	2	moi	6	poussif	1	tristesse	1
bas	1	drapeau	1	hanse	1	mois	1	pouvoir	1	trombe	1
bateau	4	du	2	heurter	1	mon	8	prendre	1	trop	1
bateauivre	1	eau	4	hideux	1	monitor	1	presque	1	trouer	1
béhémot	1	ébloui	1	hippocampe	1	monter	3	punaise	1	troupeau	1
bénir	2	échouage	1	hiver	1	montrer	1	pussent	1	tu	1
blé	1	éclair	1	homme	2	morve	1	quand	2	ultramarin	1
blème	1	éclater	1	horizon	1	mufle	1	que pron	5	un	24
bleu	4	écroulement	1	horreur	1	mur	1	que conj.	8	vacherie	1
bleuité	1	écume	1	horrible	1	mystique	1	quelquefois	1	vaste	1
blond	1	électrique	1	houle	1	nacreux	1	querelle	1	vent	1
bohus	1	embaumer	1	hystérique	1	nager	1	qui	4	ventouse	1
bon	1	en	3	île	2	nasse	1	quille	1	vers	3
bonace	1	enfant	4	illuminant	1	navrant	1	ravi	1	vert	3
bord	1	enivrant	1	immobilité	1	ne	4	récif	1	victime	1
bouchon	1	enlever	1	impassible	1	neige	1	reculer	1	viguteur	1
braise	1	énorme	1	incroyable	1	ni	2	regretter	2	vin	1
bride	1	entonnoir	1	ineffable	1	niais	1	repêcher	1	violet	2
brume	1	épais	1	infuser	1	noir	2	ressac	1	voguer	1
brun	1	équipage	1	inouï	1	noire	1	rester	1	vogueur	1
carcasse	1	escorter	1	insoucieux	1	notre	1	rêver	1	voilier	1
catarac.	1	est	1	instant	1	noyé	2	rouge	1	voir	5
ce pron.	2	et	28	ivre	2	nu	1	rougeoyant	1	volet	1
ce adj.	6	éternel	2	jaune	2	nuit	3	rouler	1	vomissure	1
cerveau	1	éther	1	je	24	ô	4	rouleur	1	votre	1
chair	1	être	6	jeté	1	océan	1	roulis	1	vouloir	2
chantant	1	Europe	2	jonc	1	oeil	5	rousseur	1	vous	1
chanteur	1	éveil	2	jour	1	oiseau	3	rut	1	vrai	1
cheveu	1	exalté	1	juillet	1	ombre	1	rutillement	1	zone	1

Tableau 3. Le vocabulaire du *Bateau ivre*

D'un gradin d'or, - parmi les cordons de soie, les gazes grises, les velours verts et les disques de cristal qui noircissent comme du bronze au soleil, - je vois la digitale s'ouvrir sur un tapis de filigranes d'argent, d'yeux et de chevelures. Des pièces d'or jaune semées sur l'agate, des piliers d'acajou supportant un dôme d'émeraudes, des bouquets de satin blanc et de fines verges de rubis entourent la rose d'eau. Tels qu'un dieu aux énormes yeux bleus et aux formes de neige, la mer et le ciel attirent aux terrasses de marbre la foule des jeunes et fortes roses.

La couleur est à la mode dans la deuxième moitié du XIX^e siècle. Les écrivains rivalisent avec les peintres, Baudelaire avec Delacroix, Zola avec Cézanne et Fromentin avec lui-même. De Théophile Gautier à Huysmans, du Parnasse au symbolisme, et du romantisme au naturalisme, la couleur survit à toutes les modes, comme chez les grands couturiers de notre époque. Cette explosion des couleurs (et des valeurs) est très visible dans la courbe de la figure 14, qui cumule les effectifs observés pour une dizaine de coloris fondamentaux.

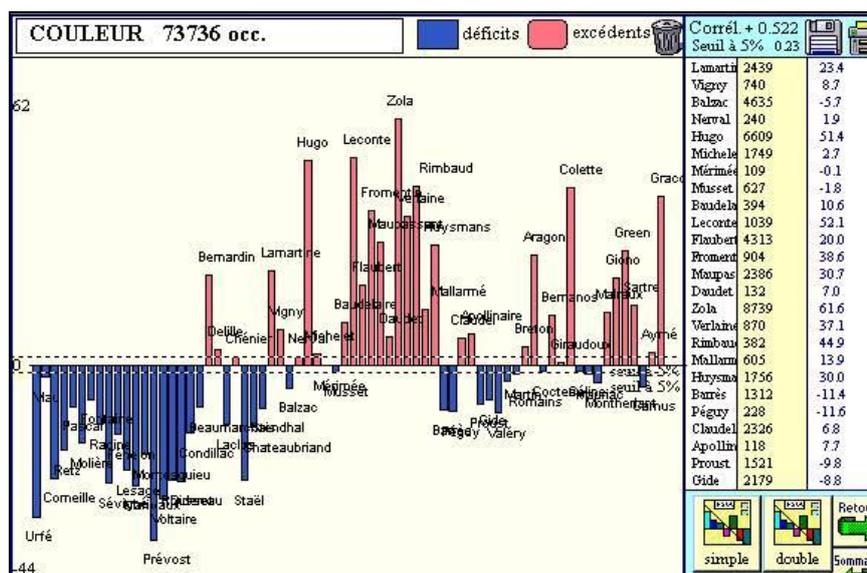


Figure 14. Explosion des couleurs au XIX^e siècle

Or Rimbaud s'installe au haut de la courbe en compagnie de Zola, de Hugo et de Leconte de Lisle. Rimbaud, qui a pris ses leçons dans les correspondances de Baudelaire, propose même une théorie des couleurs et des sons associés dans le fameux sonnet des *Voyelles*. De savants

exégètes, suivant en cela Enid Starkie²¹, estiment que dans « *A noir, E blanc, I rouge, U vert, O bleu* », l'ordre des couleurs suit les phases magiques par lesquelles passe la matière pour aboutir à l'or suprême, élixir de longue vie, dans le processus de l'alchimie. Le dernier état avant l'aboutissement est le bleu ou le violet (« *O l'oméga, rayon violet de Ses Yeux* »).

Or la prévalence du *bleu* chez Rimbaud est perceptible et observable dans nos relevés. Si Zola est le champion du *jaune* et du *rouge* (et aussi du *blanc*, à cause du *Bonheur des dames* et du *Rêve*²²), c'est Rimbaud qui l'emporte dans notre littérature pour le *vert*, le *bleu* et le *violet*. Verlaine est en tête pour le *clair*, Gracq pour le *gris*, tandis que Hugo accapare le *noir* et le *sombre*²³. On a réuni les principaux coloristes de notre littérature pour observer leur palette dans la figure 15. Hugo à droite, et Zola à gauche ont les choix les plus tranchés, tandis que Rimbaud s'affirme en bas du graphique, avec ses couleurs préférées.

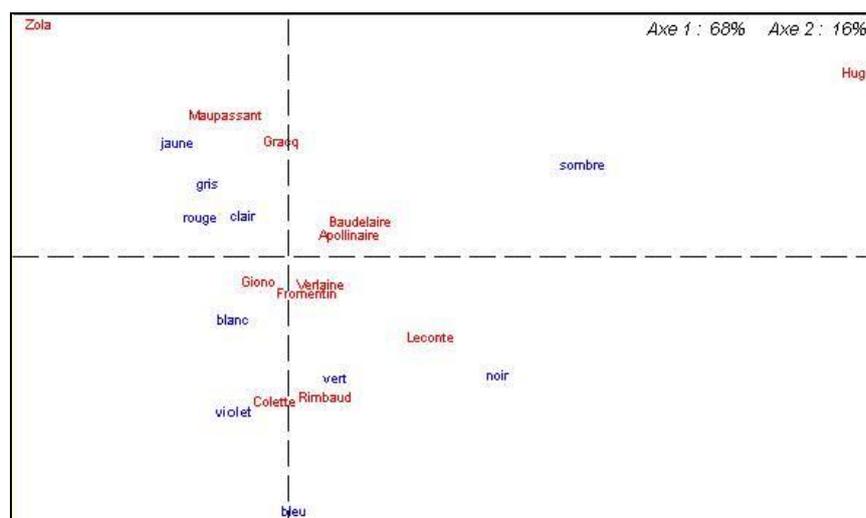


Figure 15. Analyse factorielle des couleurs chez quelques écrivains

21. Enid Starkie, *Arthur Rimbaud*, traduction de Alain Borer, Flammarion, 1982. Voir en particulier le chapitre « Alchimie et magie », p. 191-211.

22. Nos renvoyons le lecteur à notre ouvrage *Le Vocabulaire de Zola* (Slatkine, Genève, 1985), et notamment au chapitre consacré à la couleur, tome 1, p. 303.

23. *Sombre* est l'adjectif le plus spécifique de Hugo, *ombre* occupe le même rang parmi les substantifs et la paire *ombre-sombre* parmi les rimes.

En réalité, quand la concurrence est moins rude, et qu'elle se borne à Verlaine et Baudelaire, Rimbaud s'octroie la part du lion pour presque toutes les nuances, qui ont de larges excédents sous sa plume. Baudelaire apparaît bien terne dans le graphique 16 et si Verlaine colore son écriture dans les premiers recueils, les tons sont délavés par la suite. Chez Rimbaud les couleurs perdent aussi de leur intensité au cours du temps. C'est dans le premier recueil que l'excédent est le plus éclatant, alors qu'il n'est plus significatif dans les *Illuminations*, et qu'apparaît un net déficit dans la *Saison en enfer*.

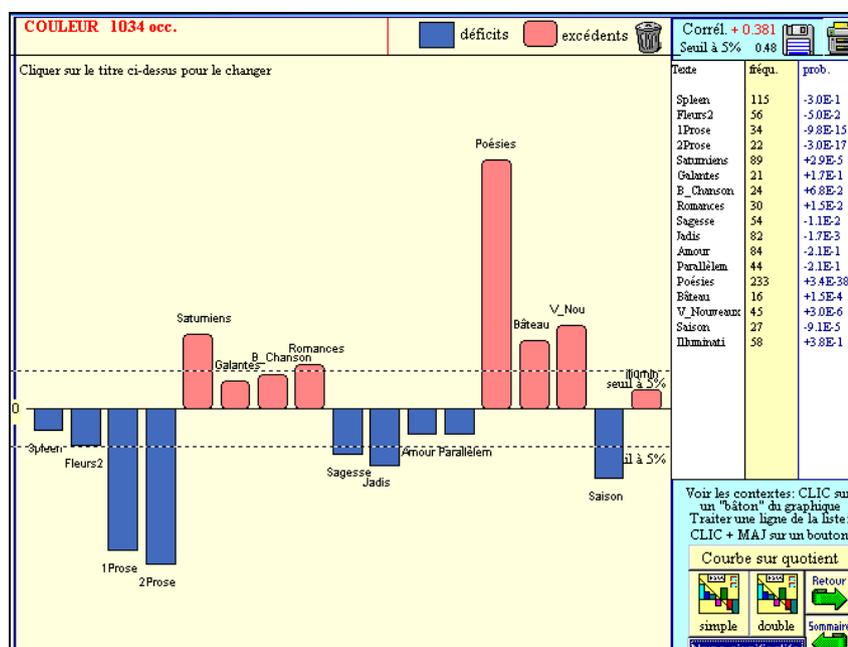


Figure 16. La courbe des couleurs chez Baudelaire, Verlaine et Rimbaud

C'est ce qui explique la position centrale des *Poésies* de Rimbaud dans l'analyse factorielle de la figure 17. Comme toutes les variétés s'y trouvent cultivées à foison, le refus de choisir entre elles aboutit à une position proche du centre de gravité. Mais la figure illustre nettement le contraste entre les couleurs sombres où se plaît Baudelaire et les teintes claires et légères dont s'entoure Verlaine. Quant aux coloris vifs, Verlaine ne les montre que dans les premiers recueils, qui sont d'inspiration parnassienne, et se rapprochent de Rimbaud, établi dans le quadrant supérieur gauche. Du jaune au rouge, et du bleu et au vert, les couleurs de l'arc-en-ciel sont l'apanage de Rimbaud.

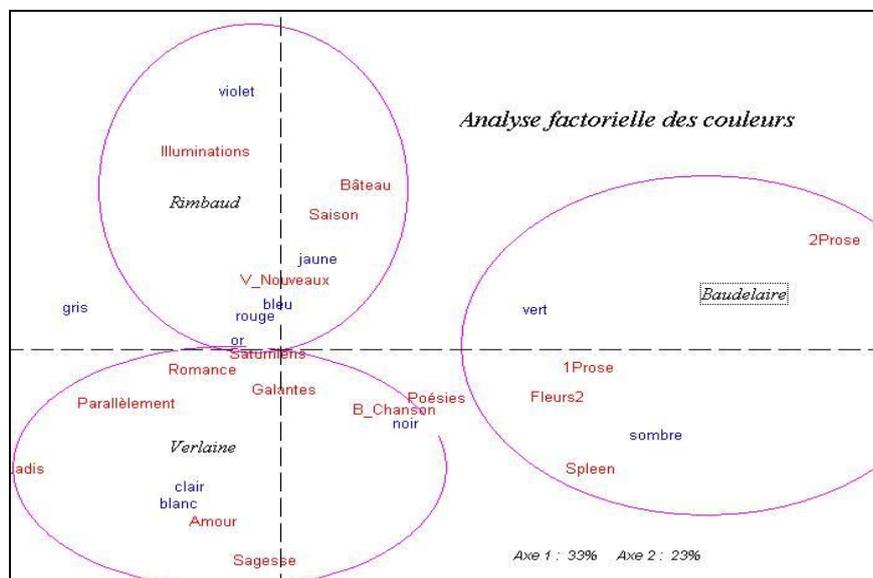


Figure 17. Analyse factorielle des couleurs (corpus Baudelaire-Verlaine- Rimbaud)

La tentation est grande de vérifier dans le texte même la théorie rimbaldivienne des voyelles. Rien n'est plus facile que de faire le décompte de chaque voyelle dans chacun des textes du corpus. Et pour faire bonne mesure la machine fera aussi le relevé de chaque consonne. Le hasard ne préside nullement aux combinaisons variables où se mêlent les deux espèces. En réduisant l'alphabet à un codage binaire où ne subsistent plus que deux codes (*v* pour toute voyelle rencontrée dans le texte, et *c* pour toute consonne), le calcul de distance permet de distinguer les trois écrivains. Mais le dosage interne des lettres ne conduit pas à des résultats qui puissent infirmer ou confirmer le sonnet des *Voyelles*. Sans doute voit-on plus de *e* et de *i* dans la prose, plus de *o* et de *u* dans les vers. Mais il s'agit de graphèmes, non de phonèmes. Si la correspondance entre les lettres et les sons reste fragile dans le cas de certaines consonnes, elle est hautement suspecte lorsqu'il s'agit des voyelles et par exemple le son *o* est perdu dans la graphie *eau*. À plus forte raison la correspondance qui pourrait éventuellement lier les sons et les couleurs échappe-t-elle à toute mesure automatique et plus encore à toute interprétation. Guiraud s'y était essayé, dans sa thèse sur Valéry, à partir de comptages manuels. Sur ce point il faut reconnaître que la machinerie n'a guère avancé, le son étant aussi rebelle que le sens à l'analyse des automates. Et c'est sur ce constat d'échec qu'il nous faut clore notre parcours.