

L'ordinateur peut-il aider à lire le théâtre ?¹

Imprimé, le texte théâtral répond à une double exigence. En tant que texte, il doit être lisible en sa totalité. En tant que genre, il comporte un certain nombre d'indications (nom du locuteur, personnages en scène, indications scéniques et autres didascalies) facilitant la représentation mentale et, à plus forte raison, scénique. Les logiciels d'analyse textuelle fonctionnant sur micro-ordinateurs étant désormais d'accès commun, on peut cependant se demander s'ils sont adaptés à cette ambiguïté du texte dramatique, et s'ils peuvent servir à mieux comprendre sa spécificité. Accessoirement, je m'interroge sur la faculté qu'ils nous offriraient de caractériser les œuvres dramatiques par comparaison entre elles.

J'essaierai de répondre à ces questions en prenant pour exemple la trilogie de Jarry : *Ubu intime*² (version primitive d'*Ubu cocu*), *Ubu roi* et *Ubu enchaîné*³, dont la vertu, en tant que textes dramatiques, n'est plus à établir⁴.

* *
*

Dans un premier temps, je ferai passer le texte d'*Ubu roi* sous le logiciel PISTES⁵. Outre les index et les concordances fournis par tout logiciel d'analyse, que peut nous enseigner un tel traitement informatique quant aux caractères généraux du vocabulaire de cette pièce ? D'abord sur la richesse lexicale. L'œuvre compte 13 229 occurrences pour 2 445 formes, ce qui donne un rapport de 5,4, peu significatif en lui-même, compte tenu de la relative brièveté du texte. Mais si, introduisant une partition, on étudie les caractères statistiques du vocabulaire acte par acte, on observe que ce même rapport nombre d'occurrences/ nombre de formes se présente ainsi : Acte I = 3,39 ; Acte II = 3,02 ; Acte III = 3,21 ; Acte IV = 3,28 ; Acte V = 3,02. On peut alors émettre l'hypothèse que si la règle (implicite) de non répétition des mots est la marque d'une écriture adulte et contrôlée, alors les actes II et V sont plus particulièrement de la main de Jarry, tandis que les autres refléteraient la création juvénile et collective du lycée de Rennes. Mais l'écart entre ces nombres est bien trop faible pour qu'on puisse en tirer une conclusion définitive. Ceci n'est qu'une piste de recherche, que des calculs plus fins portant sur les différents actes de la trilogie pourraient corroborer (à ceci près qu'*Ubu intime* n'est pas divisé en actes, et que son texte représente la moitié d'*Ubu roi*). Cependant, il est évident qu'une pièce de théâtre, par sa nature même, comporte un grand nombre de répétitions, dues aux didascalies dans leur ensemble, aux indications de rôles en particulier. Un nouveau traitement, séparant le dialogue des autres parties du discours dramatique, en fonction des personnages, me fournit les indications suivantes :

Désignation	91 formes	1094 occ.	12,02
Didascalies	463 formes	1292 occ.	2,79
Père Ubu	1524 formes	5915 occ.	3,88
Mère Ubu	701 formes	1824 occ.	2,60
Palotins	274 formes	515 occ.	1,87
Bordure	235 formes	440 occ.	1,87
Roi	138 formes	209 occ.	1,51

¹ Pré-publication dans *Analyse textuelle et nouvelles technologies*, [université Paris VIII] n° 2, 1993, pp. 13-23.

² J'ai saisi, au lecteur optique, le texte de l'édition procurée sur le manuscrit par Henri Bordillon dans : Alfred Jarry, *Ubu intime* pièce en un acte et divers inédits autour d'Ubu, Editions Folle Avoine, 1985, pp. 117-150. Il a l'avantage de restituer l'état le plus primitif du manuscrit, daté de 1894-95 par le présentateur.

³ Pour le texte d'*Ubu Roi*, nous disposons de la saisie effectuée à Nancy par l'INaLF sur l'édition des *Œuvres complètes*, t. IV, Monte-Carlo, 1948, pp. 27-92, que nous avons reconvertie manuellement selon les indications de l'édition Gallimard, coll. « Folio » (1978), la plus accessible actuellement ; en revanche, il nous a fallu saisir le texte d'*Ubu Enchaîné* à partir de la même édition (pp. 185-257).

⁴ Voir, à ce sujet, notre *Jarry dramaturge*, Nizet, 1980, 304p.

⁵ Pierre Muller, *Pistes Pour une Investigation Systématique des TExteS*, INRP, 1989.

Reine	164 formes	248 occ.	1,51
Bougrelas	227 formes	417 occ.	1,83
Tous	175 formes	341 occ.	1,94

Ce tableau appelle plusieurs remarques :

1. Le logiciel PISTES n'accepte que douze parties au maximum, ce qui oblige à regrouper certaines répliques de ce drame épique à plus de trente personnages sous le nom d'un seul d'entre eux, comme « Palotins » pour Giron, Pile et Cotice. Cette contrainte n'est finalement pas trop gênante car, inversement, la faible quantité de formes utilisées par chacun interdirait toute exploitation statistique.
2. Si l'on met à part la désignation des personnages, qui a un taux de répétitivité très élevé, on voit que les didascalies, assurément écrites par Jarry au moment où il entend livrer sa pièce au metteur en scène, Lugné-Poe, témoignent d'une certaine richesse de vocabulaire.
3. C'est bien le Père Ubu qui tient le rôle principal de la pièce. Il parle largement trois fois plus que la Mère Ubu, et presque trente fois plus que le roi. Mais le rapport des occurrences aux formes reste, chez lui, remarquablement constant, quel que soit le traitement mathématique qu'on lui fait subir. Avec un taux de 3,88, le « parler Ubu » est plus répétitif que celui des autres personnages, tel que le Roi, la Reine et même Bougrelas, drapés dans leur majesté de langage. Ce qui ne signifie pas qu'il soit dépourvu d'originalité, comme nous le verrons par la suite.
4. Ce « parler Ubu », s'il contamine la Mère Ubu, Bordure et les Palotins, constitue un groupe de sujets parlants nettement opposé à la royauté légitime (incluant le Czar Alexis).
5. Il faut noter, enfin, qu'à plusieurs reprises divers personnages (regroupés sous la rubrique « Tous ») parlent *en même temps*, comme s'ils constituaient le chœur dans la tragédie grecque, ce qui ne va pas sans poser de problèmes pour l'interprétation dramatique, et qui dénote le caractère épique de la pièce.

Par ailleurs, un rapide coup d'œil jeté au tableau de répartition des fréquences, que fournit la machine, m'indique que 58,5 % des occurrences sont des hapax, ou plus précisément qu'elles ont la fréquence 1, ce qui n'est pas tout à fait la même chose, puisque cette catégorie contient aussi bien des termes spécifiques à Jarry que ceux du langage commun qu'il n'utilise qu'une fois. Cela conforte notre intuition qu'*Ubu roi* est, sinon d'une grande richesse lexicale, terme fort controversé parmi les lexicologues, du moins d'une grande variété de vocabulaire.

L'index hiérarchique global, sans distinction des formes fonctionnelles ou lexicales, fournit de précieuses indications sur ce vocabulaire que j'ai lemmatisé manuellement pour les verbes, afin d'obtenir des données plus proches de la réalité langagière.

Le mot qui vient en tête, cela ne surprendra personne, est Ubu (561 occ.) désignant le Père Ubu (355 fois) et la Mère Ubu (178 fois) dans le dialogue et les indications scéniques, ce qui revient à marquer l'omniprésence du couple, et la prégnance de cette dénomination originale. Suit le mot « de » (381 occ.), qui tient toujours la tête dans tous les tableaux du même type, puis le verbe « être » lemmatisé (363 occ.), la conjonction « et » (287 occ.), « je » (275 occ.), « le » (268 occ.), « la » (261 occ.), le verbe « avoir » lemmatisé (241 occ.), « à » (239 occ.), « il » (190 occ.), « vous » (162 occ.). J'arrête là, au dixième rang, l'énumération des formes de plus haute fréquence. Le lecteur accoutumé à de telles données observera la présence anormale, parmi les fréquences élevées, de la première personne et des exclamatifs (« ah » = 97 occ. ; « oh » = 59 occ. ; « eh » = 53 occ.), ce qui dénote bien le dialogue théâtral, fortement ponctué par la voix. Puis il relèvera les formes « bien » (75 occ.), « fort » (23 occ.) et parmi les mots pleins les plus significatifs : « finances » (33 occ.), « merdre » (33 occ.), « mort » (30 occ.), « messieurs » (20 occ.), « sire » (20 occ.), « ours » (19 occ.), « nobles » (18 occ.), « cheval » (17 occ.), « russes » (15 occ.), « polonais » (12 occ.). À quoi s'ajoutent les formes et syntagmes du parler spécifique d'Ubu : « chandelle verte » (14 occ.), « phynances » (10

occ.), « bougre » (8 occ.) « oneilles » (8 occ. mais ce mot n'est pas l'apanage exclusif d'Ubu puisqu'un des soldats parle du « ciseau à oneilles »)...

PISTES procède, lorsqu'on le lui demande, au calcul des spécificités, c'est à dire à la comparaison de la fréquence probable d'un terme, dans chaque partie du texte, avec la fréquence réelle observée. Ce qui détermine des spécificités positives ou négatives. Je n'entrerai pas dans le détail de ce calcul hypergéométrique, me contentant des résultats affichés par la machine⁶.

Le découpage par actes met en évidence, au moyen des spécificités positives, les personnages essentiels, les formes fonctionnelles et lexicales les plus remarquables, c'est-à-dire les plus inattendues dans la totalité de l'acte. De fait, cela revient à caractériser le thème principal de l'acte, et le langage qui s'y rapporte. Ainsi, le premier acte est celui du repas au menu si étonnant et du complot. Il est dominé par la personnalité égocentrique d'Ubu, et par des formes verbales au futur. Autrement dit, rien d'autre que ce qu'une lecture non prévenue nous apprendrait ; rien que le spectateur n'ait retenu, une fois le rideau tombé, si ce n'est que l'archaïsme « estes », employé deux fois par la Mère Ubu, une fois par le Roi, disparaît à tout jamais de leur discours. L'assassinat du roi lors de la revue, à l'acte II, met en relief le rôle du jeune Bougrelas, ce qui se rapporte à son héritage symbolique et sa nécessaire vengeance. Le troisième acte est celui du verbe être, en proportion surprenante. Il montre les exactions du Père Ubu en quête d'argent et met en évidence la scène du conseil, que d'ailleurs toutes les anthologies reprennent, non sans raison. L'acte IV est bien celui de la bataille. J'observe seulement qu'Ubu, le Roi, la Reine, Bordure, les nobles, la chandelle verte et madame y sont anormalement absents, comme si la guerre était affaire de soldats et non de princes ! L'acte V voit la réapparition de la Mère Ubu ; il met en relief le lexique propre à la navigation, dans la scène ultime.

En somme, pour hâtive qu'elle soit, cette analyse souligne les scènes marquantes de l'œuvre. Elle pourrait confirmer les propos de l'auteur qui prétendait avoir cousu entre elles des scènes élaborées huit ans auparavant par des potaches.

Changeant de partition, je me suis intéressé au langage de chaque individu, selon la même méthode. Et d'abord aux didascalies, sans quoi nous n'aurions pas les signes extérieurs du dialogue théâtral. Il apparaît que celles-ci constituent un véritable langage, ne désignant pas seulement les personnages, les lieux et les objets, qualifiant aussi les actions, les mouvements. On notera en particulier l'emploi exclusif du verbe « déchirer », au sujet d'êtres humains. Inversement, le « parler Ubu » contamine leur auteur, puisqu'on y lit cette phrase : « Il lui découde la *boudouille* d'un terrible coup d'épée ».

J'ai déjà esquissé quelques réflexions sur le parler d'Ubu, le nombre inespéré de ses tournures exclamatives, la haute fréquence de la conjonction de coordination (147 occ. sur 281, probabilité positive 1.5E-02), signe d'une tournure enfantine : « Et puis après ? » ; son emploi immodéré des formes de la première personne (fréquence observée : 376 ; fréquence théorique : 266 ; probabilité positive : 2.5E-21), signe incontestable de son autoritarisme. Si je m'en tiens aux seuls mots pleins, je relève la liste suivante parmi les spécificités positives : finances (28 sur 33 occ.), phynances (9 sur 10 occ.), merdre (25 sur 33 occ.), chandelle verte (dont il fait un usage exclusif), maître, trappe, oneilles, physique, corne, peur (5 occ. sur 5), biens, caisse, poche, bâton, cornegidouille, moitié (4 sur 4 occ.), sabre, salopins, torsion, tour et l'adjectif « petit » qui lui est quasi consubstantiel (12 occ. sur 13, avec une probabilité positive de 6.7E-04). Voilà bien, rapidement esquissé, tout l'univers verbal fait de néologismes, d'insultes et de jurons, de sentences professorales et d'expressions enfantines, constitutif de l'ici présent maître des phynances.

Celui de Madame sa fumelle, plus bref, s'établit ainsi : Monsieur, fais, tué, beau, faut, grâce, jours, tes, veau, avoir, femme, pauvre, parti, bourrique, écoutez, pierre, rois, trésor, pardonner.

⁶. Pour le détail de ce calcul, voir le manuel de Pierre Muller cité ci-dessus, et surtout la thèse de Pierre Lafon : *Dépouillement et statistiques en lexicométrie*, Paris-Genève, Champion-Slatkine.

Il résume ainsi l'essentiel de ses préoccupations, centrées sur la gidouille du Père Ubu, qualifié de « gros P.U., gros pantin, gros polichinelles, mon gros bonhomme » (5 occ. sur 7), où la première personne est singulièrement absente (70 occ. contre 376 pour le Père Ubu). Quant à celui des Palotins, je ne relèverai que deux formes parmi les plus caractéristiques : « monsieur » (3 occ. sur 7) et leur grognement exclusif, « hon » (6 sur 6).

On pourrait poursuivre cet examen, en soulignant ce qui caractérise, sur le plan purement statistique, le langage des uns et des autres. Le parler du Roi est celui d'un insouciant où se signalent l'intensif « fort », le comparatif « comme », l'archaïque « estes », les mots « revue, demain, donnerai, fasse, mirliton ». La Reine témoigne de sa croyance en Dieu, de son souci de la famille, de ses craintes (hélas = 2 occ. sur 5). Quant à Bougrellas, s'il reprend en partie le même vocabulaire, il le complète par l'appel aux amis, au père et à la mère, à la vengeance enfin. Le vocabulaire de Bordure est bien celui d'un militaire (tuer, coup, épée, hommes, peuple...) c'est pourquoi je n'en dirai rien de plus.

Les outils informatiques nous permettraient, à partir de ces observations externes, d'analyser de plus près l'usage, dans la pièce, de chacun des termes signalés, de repérer les champs lexicaux dominants, les différents niveaux de vocabulaire, singulièrement mêlés dans ce discours dramatique. Je songe au vocabulaire de la nourriture, à celui des combats, au style noble, archaïque, aux tournures familières, aux injures. Comme cela a été traité jadis par Michel Arrivé dans ses *Langages de Jarry*⁷, je n'y reviendrai pas, puisque je ne pourrais qu'étayer ses remarques par des listes exhaustives, donc lassantes.

En revanche, il sera plus pertinent de soumettre la trilogie ubuesque à l'analyse automatique. Changeant de produit, je me suis alors servi du logiciel HYPERBASE, procuré par Étienne Brunet⁸, afin d'en explorer les fonctionnalités. Non que PISTES, dont les capacités de traitement sont limitées par la mémoire vive des ordinateurs au standard IBM ne puisse absorber un ensemble de 33 555 occurrences concentré sur 4 511 formes, soit un rapport de 7,44, qui se restreint à 6,15 si l'on exclut la désignation des personnages en scène, comme je l'ai fait pour éviter ce brouillage propre au texte dramatique. Cette version d'HYPERBASE est susceptible d'absorber, sur le standard Macintosh, un corpus d'un millier de pages. Elle est particulièrement conviviale et ne nécessite qu'un codage très sommaire (en revanche, le traitement des sous-parties y est plus délicat).

La première originalité d'HYPERBASE est d'offrir une comparaison (ramenée à l'écart réduit) des formes du corpus choisi avec celles de l'ensemble du Trésor de la Langue Française (TLF) de Nancy, lui-même considéré comme un corpus littéraire. Un regard rapide sur les listes issues de l'ordinateur confirme l'impression que donne *Ubu roi* : c'est bien un discours dramatique, comme le montre la surabondance des points d'exclamation (1021, écart réduit 62,2) et d'interrogation (238, écart réduit 10,7), des exclamatifs : O (33, ER=41,5), oh (64, ER=19,1), voyez (64, ER=32), vive (47, ER=28), ah (112, ER=26,8), ha (4, ER= 6,9), eh (57, ER=17,9), voilà (72, ER=16,4) etc. Les marques particulières du dialogue (497 je, ER=12,5 ; 107 ma, ER=9 ; 358 nous, ER=19,9 ; 105 notre, ER=16 ; 395 vous, ER=18,5 ; 154 tu, ER=13,7 ; 60 te, ER=9,7 ; 44 toi, ER=7,8) y sont statistiquement supérieures à la moyenne du TLF.

De même, il n'est pas inutile de s'assurer, chiffres à l'appui, que la trilogie ubuesque emploie un certain nombre de termes de façon significativement excédentaire par rapport au TLF. Pour ne pas ennuyer le lecteur par une énumération fastidieuse, citons, dans l'ordre décroissant les formes dont l'écart réduit va de 64,9 à 10 : finances, chandelle verte, monsieur, livres, ours, tonneau, polonais, sire, esclave, roi, caporal, impôts, tapage, corne, russes, sauve, vais, veau, valise, messieurs, nobles, prison, oncle, tiens, dragons, gueux, bien, caisses, chaussures, sonne, diligence, secours, avons, indigne, esclavage, madame, respectable, fossé,

⁷. Voir : Michel Arrivé, *Les Langages de Jarry*, Paris, Klincksieck, 1972, 384p.

⁸. Étienne Brunet, *Hyperbase* (version réduite), logiciel pour le traitement documentaire et statistique des corpus textuels, Nice, 1991.

cul, coffre, caisse. Ce joyeux fourre-tout montre bien la large tessiture du langage de Jarry, qui va du noble au trivial, emploie un personnel dramatique à la fois noble et bourgeois et désigne aussi bien les objets du trône et du combat que de la boutique. Inversement, ce sont surtout des formes fonctionnelles qui paraissent déficitaires par rapport au même ensemble témoin : l', elle, était, avait, son, un/e, qu', de/s, d', dans, ses, qui, se, dont, du, comme, la, lui [...] et, parmi les formes lexicales : cœur, esprit, âme, yeux, amour, corps. En somme, si l'on ne regarde que les hautes fréquences, le langage théâtral d'Ubu se démarque de la langue littéraire des XIX^e-XX^e siècles par un certain nombre de termes caractéristiques de la forme dialoguée, qui constituent un univers bien particulier où les hautes fonctions de l'individu, même quand elles sont nommées, ne sont là qu'en creux, pour marquer une sorte de dérision par rapport à ce qui constitue, habituellement, la grande littérature.

Si l'on observe la trilogie en elle-même, et non plus par rapport à un corpus externe, on constate tout d'abord que 460 formes sont communes à chacune des parties, soit un dixième de l'ensemble, ce qui assure sa cohérence. Ainsi des termes que l'on croirait appartenir au seul *Ubu roi* se trouvent dans les deux autres pièces : Aragon, cornegidouille, décerveler, épouse, extraction, finance, foutre, gidouille, grotesque, oneilles, palotins, phynances, sagouin, tudez... Il est clair qu'en les employant même négativement, le dramaturge Jarry a voulu marquer la continuité de l'une à l'autre, confirmant la thèse d'une geste potachique dont il n'aurait été que le pieux et honnête scribe. L'analyse factorielle (dont je ne puis songer à détailler ici la procédure ni reproduire les graphiques) d'une partie de ce vocabulaire commun, aussi bien que des fréquences supérieures à 50 confirme qu'*Ubu enchaîné* est bien, selon le mot de Jarry, « la contrepartie d'*Ubu roi* », c'est-à-dire son symétrique inverse, tandis qu'*Ubu intime* se trouve dans une nébuleuse située dans le quadrant supérieur droit. Peut-on pousser davantage la réflexion pour en déduire que telle pièce daterait bien de 1888, telle autre de 1895, telle autre enfin de 1901 ? Le délai me paraît bien trop court pour qu'on puisse s'avancer sur un tel terrain, d'autant plus que Jarry, comme on vient de le voir, s'est évertué à maintenir une cohérence d'ensemble en raboutant les fragments de la geste pour diverses représentations théâtrales. La graphie différente des exclamatifs (o/oh, hé/eh) n'est pas déterminante puisqu'elle est davantage liée aux usages typographiques qu'à la pratique de Jarry, d'autant moins contrôlable que nous ne pouvons la vérifier sur les manuscrits originaux. La seule chose que l'on puisse montrer, par le calcul de l'écart réduit (HYPERBASE) ou celui des spécificités (PISTES), est qu'*Ubu intime* dénote le vocabulaire et les tics professoraux, qu'*Ubu enchaîné* explore le champ sémantique de la servilité. Quant à *Ubu roi*, les deux graphiques ci-joints en soulignent les traits essentiels. On aborde ici les champs lexicaux propres à chaque pièce.

Je me tournerai alors vers un second type d'outils informatique, la Banque de données d'histoire littéraire (BDHL) dont l'un des objectifs, qui n'est pas le moins hardi et par conséquent le moins controversé, consiste à dégager la thématique des œuvres retenues par l'histoire.

Dans cette banque, *Ubu roi* est caractérisé par les concepts-clés suivants, dans l'ordre alphabétique : ambition, argent, avarice, guerre, lâcheté, laideur, tyrannie, volonté de puissance (à quoi on pourrait ajouter : château, cupidité, or, trésor, ce qui ne fait que préciser la liste précédente). On objectera, à juste titre, qu'une telle énumération peut s'appliquer aussi bien à une tragédie qu'à une bouffonnerie, ce qui est exactement le but recherché, car, à la limite, nous voulons ainsi mettre sur le même plan des œuvres de nature différente traitant du même contenu. En somme, que les mêmes concepts s'appliquent parfaitement à telle œuvre de Shakespeare ou de Corneille, ce serait la preuve de leur cousinage.

S'agissant de littérature française, j'ai cherché, parmi les œuvres dramatiques enregistrées dans cette banque entre 1850 et 1900 celles qui contenaient, en tout ou en partie, les mêmes concepts. Voici ce que j'y ai trouvé :

1882 – *Les Corbeaux* (Becque) : argent, avarice, capitalisme, célibat, criminel, escroquerie, évasion, famille, homme de loi, mariage, mésalliance, militaire, misère, mort, orphelin, raison, richesse, sacrifice, travail.

1890 – *Axel* (Villiers) : affrontement, bonheur, château, connaissance, cupidité, dieu, emprisonnement, héritage, homicide, idéalisme, mort, or, passion, pureté, religion, retraite, rêve, suicide, trésor.

1890 – *Tête d'or* (Claudel) : aventure, connaissance, courage, défaite, fidélité, filiation, gloire, guerre, jeunesse, monarchie, mort, mystère, noblesse, occupation, révolte, roi, ruine, volonté de puissance.

Entre un drame naturaliste et deux drames symbolistes, l'œuvre de Jarry n'est pas en mauvaise compagnie. On comprend quelles en sont les racines spirituelles, du moins pour les spectateurs de l'Œuvre, sinon pour l'auteur, qui n'a pas laissé de confidences sur ses lectures à l'âge de 17 ans. Mais on voit d'emblée en quoi elle en diffère, par son schématisme, par l'extrême pureté de l'intrigue, si j'ose dire. Il n'y a pas là de conflit de famille, d'amours plus ou moins contrariées. Tout y est subsumé à l'état de concept, comme l'ont fort bien vu, par la suite, Ghelderode et Ionesco. Nous sommes effectivement en présence d'une synthèse et d'un dépassement des esthétiques naturaliste et symboliste, comme je l'indiquais en 1967 dans mon *Étude sur le théâtre dada et surréaliste*, ouvrant la voie à une nouvelle dramaturgie. Mais le plus étonnant est que, si *Axel* fut joué, non sans difficultés ni controverses en 1894, il fallut attendre plus de cinquante ans pour que *Tête d'or* fût porté à la scène, tandis qu'*Ubu* hantait depuis longtemps toutes les consciences. De sorte qu'on peut se demander, à bon droit, si Jarry n'est pas le père de Claudel – sur le plan dramatique s'entend.

* *
*

J'ai bien conscience de n'avoir pas fait rendre à l'ordinateur tous les calculs extrêmement sophistiqués dont il est capable. Je ne me suis pas servi, non plus, des concordances et des contextes que la machine offre à foison à qui veut les cueillir, dont nous savons quelle précieuse aide ils apportent à l'étude des champs sémantiques *présents dans le texte* et qui, regroupés avec un minimum de perspicacité permettent de dire rapidement « de quoi ça parle ». C'est d'ailleurs ce que soulignent les spécificités (ou l'écart réduit) de chaque partie, pièce ou actes. Si de tels calculs nous semblent souvent pointer des évidences, que la simple lecture cursive a déjà perçues, ils peuvent servir à caractériser autre chose, comme la voix de chaque rôle ou le langage particulier des didascalies. Ceci au prix d'un codage minimum, dont le rendement me semble des plus intéressant.