



HAL
open science

Le TLF : un trésor est caché dedans

Étienne Brunet

► **To cite this version:**

Étienne Brunet. Le TLF : un trésor est caché dedans. Annales de la Faculté des Lettres et Sciences Humaines de Nice, 1977, 29, pp.261-273. hal-01424814

HAL Id: hal-01424814

<https://hal.science/hal-01424814>

Submitted on 9 Jan 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.

Etienne BRUNET

LE T.L.F. : UN TRÉSOR EST CACHÉ DEDANS

Le T.L.F. met sous presse son cinquième tome. On sent que le rythme de parution s'accélère et l'on entrevoit à l'horizon la terre promise des lexicographes, cette lettre Z où « Zut » signifie « Ouf » et où, la tâche achevée, on savoure, en gardant le noyau dans la bouche, ces mots divins : le « zizi », le « zombi », le « zygoma ». A mesure que l'horizon se rapproche la colonne grossit : une équipe parisienne groupée autour d'A. Rey participe désormais à la rédaction et Julia Kristeva figure parmi les nouveaux collaborateurs. Il est loin le temps des pionniers, l'époque héroïque où, il y a presque 20 ans, cette 2e DB pacifique partait, à rebours, de Strasbourg. Et l'on s'inquiète de voir si proche le terme de l'entreprise. Qu'advient-il alors de cette puissance logistique, de ces chars sophistiqués, de cette concentration de troupes et d'experts? Quand les Immortels font un dictionnaire, ils ménagent mieux leurs forces.

Qu'on se rassure pourtant : point n'est besoin des tricheries nocturnes de Pénélope, l'avenir du T.L.F. est tout encombré de tâches et de projets. Et l'impression du visiteur est celle d'une entreprise en expansion, qui exporte et réinvestit. Son directeur, le recteur IMBS, aime à dire qu'une entreprise réussit avec du savoir, du savoir-faire, et du faire savoir. Longtemps, en effet, le Trésor s'est soucié d'amasser le savoir dans ses rayons, et de mettre en œuvre les méthodes les plus rationnelles et les plus modernes du savoir-faire. Restait la dernière obligation : faire savoir, transmettre le savoir, partager le trésor.

C'est chose faite aujourd'hui : les portes sont largement ouvertes et le trésor est à la disposition des fidèles et des badauds.

Il ne s'agit pas seulement des produits livrés à la vente, que ce soient les quatre premiers tomes du *Dictionnaire* ou les *Tables de Fréquences du vocabulaire littéraire des 19e et 20e siècles*¹. Ces produits finis et raffinés ne sont que la quintessence d'un matériau surabondant dont l'exploitation

1. *Dictionnaire des fréquences. Table alphabétique* 1971, 4 vol XXII-2235 pp. *Table des fréquences décroissances* 1971, XIII — 577 pp. *Table des variations de fréquences* 1973, VIII — 451 pp. *Table de répartition des homographes* 1969, XV — 97 pp. Didier, 15, rue Cujas. Paris.

est loin d'être épuisée. On dit même que la rédaction du *Dictionnaire* ne retient guère que deux pour cent des documents disponibles. C'est donc une terre presque vierge qu'on livre à l'appétit des chercheurs.

Vierge, mais non impénétrable. Le T.L.F. propose des guides, et notamment un service de documentation conçu primitivement pour la rédaction du dictionnaire et spécialisé dans les divers domaines de la linguistique française. Le développement de ce service a été si important qu'un système automatique de consultation a dû être mis au point, à l'aide de descripteurs disposés en arborescence et reliés horizontalement aux mots-clés synonymes et verticalement aux étages immédiats de la hiérarchie, génériques ou spécifiques². Il s'agit là d'une véritable banque de données bibliographiques, qui vient heureusement combler une lacune du *Bulletin signalétique* du C.N.R.S., plus riche en linguistique générale que française. Une logique simple d'interrogation mettant en œuvre les opérateurs « et », « ou » et « sauf » permet, par le jeu des parenthèses, de réduire ou d'élargir à volonté le champ de la recherche. A cela s'ajoute un guide bibliographique périodique, le *Bulletin Analytique de Linguistique Française*, qui depuis 1969, à raison de cinq ou six numéros annuels, consacre ses notices analytiques aux plus récentes publications sur la langue française³. La compatibilité est naturellement assurée entre le classement du B.A.L.F. et celui du *Thesaurus de Linguistique Française*, ces deux outils correspondant à des besoins complémentaires, puisque le chercheur doit tout à la fois être tenu au courant de ce qu'on publie de nouveau dans sa discipline et être en mesure de recevoir une documentation rétrospective sur un point particulier⁴. Pourvu d'un modèle par le *Bulletin Signalétique*, fort de l'expérience de ses informaticiens et de ses documentalistes et bénéficiant d'une masse documentaire unique en France, le T.L.F. ne peut pas ne pas être un carrefour où les chercheurs viendront, près de la Colline Inspirée, consulter la table d'orientation de la Linguistique Françaises.

Cette infrastructure documentaire était nécessaire à la rédaction du dictionnaire mais elle garde une valeur en elle-même, comme ces routes

2. On trouvera la description de ce système dans le supplément à la revue *Le Français Moderne*, fascicule 2, année 75, XII — 58 pp. Puisque ce système en est encore à la période des essais et des perfectionnements pourquoi ne pas prévoir la possibilité de traiter non plus seulement la réduction des ouvrages en une chaîne de descripteurs, mais de véritables résumés écrits en langage naturel par l'auteur. Certes, cela suppose que les articles et ouvrages recensés soient accompagnés de leur résumé, ce qui n'est pas le cas dans les disciplines littéraires. Mais cet usage est maintenant établi dans les milieux scientifiques et tend à s'étendre aux autres domaines. Ainsi les thèses — même littéraires — sont annoncées désormais par un résumé obligatoire à la charge de l'auteur. Il suffirait que la direction des revues linguistiques impose cette exigence et le traitement automatique en serait grandement amélioré. Le système documentaire du Centre d'Études et de Recherches de la Gaude fonctionne sur ces principes, à la grande satisfaction des usagers d'IBM Europe.

3. le B.A.L.F. est disponible à la Librairie Didier, 15 rue Cujas, Paris.

4. La documentation automatique est assurée par un ensemble de programmes connu sous le nom de MISTRAL et mis au point par la Compagnie CII.

5. Ajoutons pour mémoire que l'accueil réservé aux pèlerins de la science est d'une qualité rare, puisque le gîte est assuré au sein même du sanctuaire.

ou ces ponts qu'on a construits pour acheminer les matériaux des palais et des cathédrales et qui ont rendu bien d'autres services. C'est le cas aussi des matériaux même élaborés par le T.L.F., index, concordances, tables de fréquences, états de groupes binaires, listes de toutes sortes, énorme masse d'informations originales enfouies dans les profondeurs de l'iceberg dont le dictionnaire n'est que la partie visible. Or ces profondeurs sont accessibles : on peut consulter sur place, à Nancy, les concordances (avec un contexte de trois lignes) de tous les mots sémantiques — ou mots pleins — du fonds littéraire des 19^e et 20^e siècles. Quand un environnement plus large est souhaité, les fiches-textes associées aux Index⁶ restituent un contexte de dix-huit lignes. D'autres matériaux moins connus intéressent lexicologues et lexicographes, notamment ceux qui concernent la nomenclature. Ainsi l'étude de la suffixation peut s'appuyer sur des dictionnaires inverses (T.L.F., Petit Larousse, Petit Robert). Les études morphologiques feront leur pâture du dictionnaire des graphies, les lexicographes enfin apprécieront la « nomenclature des nomenclatures » ou liste comparative des entrées retenues par les quinze principaux dictionnaires ou inventaires, du *Dictionnaire de l'Académie* au *Trésor*, en passant par le *Littre*, le *D. G.* et le *Robert*. L'association même des mots n'échappe plus au traitement électronique, qu'il s'agisse d'une fusion lexicale (Index des mots composés dans le Corpus du XIX^e et XX^e siècles), d'une relation d'antonyme ou de synonyme (liste des adjectifs reliés par « ou ») ou de contiguïté dans la chaîne parlée, soit directe (séquences des mots grammaticaux), soit indirecte (groupes binaires). Les groupes binaires — dont l'élaboration est en cours — offrent un intérêt particulier pour le traducteur et plus généralement le linguiste, ils mettent en évidence les associations préférentielles que les mots sémantiques établissent entre eux et dont la variété est très inférieure à ce qu'elle serait dans une combinatoire libre⁷. On imagine les promesses de cette voie ouverte à la statistique, lorsque dans les groupes binaires on aura distingué la catégorie grammaticale des éléments rapprochés, ou les tranches chronologiques, ou les genres, ou les auteurs. Tout le jeu du langage peut être ainsi mesuré depuis les contraintes les plus fortes qui créent le cliché, jusqu'à la liberté folle des associations surréalistes. Et l'on verra s'opposer les auteurs, les genres ou les époques, non plus seulement par l'étendue de leur vocabulaire, mais par leur richesse associative, ce qui représente un pas supplémentaire vers la Poétique et la Stylistique.

6. Les index sont souvent spécialisés : index d'auteurs, de périodes, de genres. Ils peuvent être constitués spécialement sur demande. Certains portent sur un corpus particulier comme l'index cumulatif d'usuels de phonétique et d'orthographe, l'index encyclopédique (ouvrages de vulgarisation scientifique), l'index technique (dictionnaires techniques).

7. Il y a groupe binaire lorsque deux mots se trouvent contigus une fois les mots grammaticaux effacés. (Mais les signes de ponctuation sont maintenus et interdisent l'association).

En attendant que le relevé des groupes soit achevé, et que le dépouillement des textes soit complètement assuré⁸, une double question se pose. Comment se procurer ces documents et comment les utiliser?

Tant que le T.L.F. s'est soucié de façon prioritaire et presque absolue de rédiger son dictionnaire, les sous-produits de l'entreprise, sans être jetés au rebut ni relégués au secret, n'ont cependant pas été exploités à fond. Il fallait éviter la dispersion des efforts et laisser quelque tâche noble aux générations suivantes. Mais lorsque les matériaux nécessaires à l'objectif premier ont été emmagasinés et que la rédaction du dictionnaire a pu se poursuivre sans apport nouveau, le T.L.F. a dû s'interroger sinon sur son existence du moins sur son orientation. Et, de même que l'Académie Française n'a pas cessé d'exister le jour où parut enfin la première édition de son dictionnaire, de même le T.L.F. va-t-il renaître de son œuvre comme le phénix de ses cendres. Et ce sera désormais l'Institut de Linguistique Française. Il n'appartient pas à l'auteur de ces lignes de dévoiler l'organisation future de cet organisme. Seul le principe fondamental peut être ici souligné : le trésor deviendrait une banque. Au lieu d'entasser l'or sur place, on ferait circuler la monnaie au gré des besoins et des demandes. Et déjà sur tout le territoire français des équipes de linguistes coordonnent leurs programmes de recherche et établissent les liaisons d'un réseau cohérent où les données, les moyens et les hommes communiqueront entre eux. Le modèle financier ne suffit plus : il cède la place à l'organisation industrielle, car il ne s'agit pas seulement d'instituer une banque de données mais de répartir la tâche à des équipes spécialisées qui des données brutes aboutissent par étapes au produit fini. Ainsi les lexicographes passeraient le flambeau aux statisticiens et les grammairiens assureraient le dernier relais.

Avec l'avènement de l'informatique et le développement des télécommunications point n'est besoin en effet de concentrer les hommes Rien n'est plus facile que de transmettre une bande magnétique ou un programme. Et cela même pourrait être évité si des lignes spéciales, comme celles du réseau CADUCEE, relient entre elles les équipes linguistiques de l'hexagone, comme elles relient déjà les équipes scientifiques de la France entière, qui travaillent souvent à distance sans avoir sur place ni les données ni les moyens de calcul. Or, il se trouve que les bandes magnétiques du T.L.F. sont de lecture aisée et d'exploitation facile sur les matériels de la C.I.I. dont sont dotées la plupart des universités françaises¹¹. Cela suppose toutefois que l'utilisateur connaisse le contenu précis de chaque fichier

8. On en est à la lettre E.

9. Le corpus du XVIIIe siècle est actuellement disponible. Il comprend 600 textes (soit 30 millions d'occurrences) répartis par tranches chronologiques (de 5 ans). Cf. le supplément à la revue *Le Français Moderne*, p. 37, fascicule 1, 1974.

10. Le sigle T.L.F. pourrait alors avoir le sens de Télé Linguistique Française.

11. L'exploitation sur un matériel IBM est également possible puisque le code adopté est le même. Les résultats qui suivent ont d'ailleurs été obtenus à l'aide d'ordinateurs IBM.

et toutes les indications techniques qui conditionnent son utilisation. La mise en place d'une véritable banque de données passe d'abord par l'inventaire du trésor, la standardisation, la mise à jour et la vulgarisation d'une bibliothèque de programmes et de fichiers. On peut imaginer ensuite la fusion de tous ces fichiers en ce que les informations appellent une « base de données », c'est à dire un fichier unique où chaque article est une structure complexe pourvue de liaisons multiples qui permettent des interrogations variées¹². La réalisation d'une telle base de données est cependant difficile parce que les fichiers existants manquent parfois de cohérence ayant été créés selon les circonstances et les besoins, parce qu'il est malaisé également de prévoir les recherches futures et les questions virtuelles, enfin parce que la matière même du langage est rebelle aux classifications simplistes et aux liaisons univoques.

Le second problème est pragmatique aussi : que faire des documents obtenus? Posée ainsi la question a peu de sens : à quoi bon obtenir des données dont on n'a que faire? Au chercheur de savoir où il va et de mesurer la masse de données dont il a besoin. Le balzacien qui demande la concordance complète de son auteur doit savoir qu'il lui faut une camionnette. Or jusqu'ici l'utilisation des documents a été surtout d'ordre analytique et lexicographique, le but étant de concentrer sur tel ou tel mot le maximum d'informations. Le temps est venu de la synthèse et trois perspectives principales s'imposent dans l'immédiat.

La première suit le cours du temps. Comment la langue évolue-t-elle et quels indices de cette évolution trouve-t-on dans les variations de fréquences relevées à l'intérieur du corpus des XIXe et XXe siècles? On sait que quinze tranches chronologiques ont été distinguées de 1789 à 1964 et que les fréquences par tranches ont été publiées dans la troisième partie du *Dictionnaire des Fréquences*¹³. On prendra garde toutefois que les fréquences y sont appelées improprement relatives et qu'il s'agit en fait de fréquences réelles dont la comparaison suppose une pondération préalable. Car les tranches n'ont pas exactement le même poids et l'on constate 50 de différence de la plus petite à la plus étendue. En tenant compte du volume respectif de chacune, on peut dresser la courbe chronologique de chaque mot ou mieux d'un champ sémantique, d'une catégorie grammaticale, des suffixes, des temps verbaux, etc... A travers le discours c'est aussi l'histoire des mentalités, l'évolution des sensibilités qui devrait apparaître. Ainsi les auteurs du T.L.F. ont eux-mêmes souligné le déclin du mot « âme ». En revanche un simple coefficient de variation (ou mieux un coefficient de Spearman) dénonce le progrès facilement explicable du mot « allemand ». A qui étudie la notion de « trahison », l'histoire rend compte des accès de fièvre qui agitent les périodes 1893-1907 (C'est l'Affaire Dreyfus) ou 1946-

12. Sur le matériel C.I.I. la consultation de bases de données s'opère par le truchement d'un système appelé SOCRATE, peut-être en souvenir du philosophe qui savait si bien poser les questions.

13. *Table des variations de fréquences*, 1973, VIII - 451 pp., Didier, Paris.

1964. La guerre de 14-18 semble avoir été plus loyale que la dernière et la trahison y est plus rare, du moins apparaît-elle moins dans l'écho des événements qui se répercutent, parfois à retardement, dans la littérature. Par contre le tragique est également partagé entre les deux guerres.

TABLEAU I — VARIATIONS DE FRÉQUENCES DE 1880 à 1964

Tranche I	1880	1893	1908	1919	1927	1933	1938	1946	TOTAL
	1892	1907	1918	1926	1932	1937	1945	1964	
la trahison (trahir + trahi + trahison)									
fréquence réelle	228	473	331	322	255	351	444	584	2988
fréquence relative	2092	3877	3123	2176	2525	3026	3041	3842	2988
le tragique (tragédie + tragique)									
fréquence réelle	171	298	364	353	265	311	517	273	2552
fréquence relative	1569	2443	3434	2385	2624	2681	3541	1796	2552

Reste à démêler les rapports complexes de l'histoire, de la littérature et de la langue. Or c'est en déroulant le temps qu'on peut saisir au mieux le croisement des fils de l'écheveau. En le déroulant assez vite l'ordinateur, par une synthèse rapide assez semblable à l'effet stroboscopique, permet de dégager les lignes de force, les tendances contrariées ou conjuguées qui animent les mots, dont certains apparaissent et progressent quand les autres régressent ou disparaissent. Nul doute que *l'Histoire de la Langue Française*, dont M.G. Antoine entreprend la continuation, tirera profit des données du T.L.F. Et l'on se prend à regretter que Ferdinand Brunot et spécialement Charles Bruneau, pour l'époque romantique et l'époque réaliste, n'aient pas eu ces documents à leur disposition. La consultation du fonds Brunot, dont le T.L.F. a reçu l'héritage, est émouvante à cet égard : toutes ces notes manuscrites, presque calligraphiées, ne comprennent souvent qu'une référence ou une datation, c'est dire à quel point l'illustre historien de la langue eût apprécié nos index électroniques et nos relevés de fréquence.

La seconde voie de recherche conduit du côté des genres littéraires. On sait que le corpus des XIX^e et XX^e siècles a été divisé en trois sous-ensembles : Prose, Prose poétique et Vers¹⁴. C'est l'occasion de vérifier

¹⁴. On y trouve aussi une division selon la personne dominante : Soliloque (1^{re} personne) Dialogue (deuxième) Reste (troisième), l'appartenance à l'un ou l'autre de ces sous-ensembles étant reconnue par la machine d'après les pourcentages respectifs des pronoms personnels.

si les barrières qui distinguent les genres ont été abolies après que Hugo eut mis un bonnet rouge au dictionnaire On verra d'après la liste ci-dessous qu'il n'en est rien. Certes la poésie au niveau du vocabulaire fondamental — c'est-à-dire des fréquences moyennes — ne distingue pas entre vers et prose poétique. Et les deux sous-ensembles se recoupent en grande partie. Parmi les 240 mots les plus caractéristiques de l'ensemble en vers les deux tiers le sont également en prose poétique, avec des écarts réduits toutefois un peu moins élevés. Les mots qui semblent appartenir en propre à la prose poétique appartiennent en fait à certains auteurs, surtout Claudel et Péguy, trop bien représentés dans l'échantillonnage du T.L.F.¹⁵ On y reconnaît une inspiration religieuse qui n'est pas commune à tous les auteurs retenus dans ce sous-ensemble, non plus qu'à ceux qui alimentent le corpus en vers. Les mots *éternel*, *espérance*, *création*, *créature*, *église*, *chrétien*, *salut*, *gloire*, *sauver*, ont des connotations chrétiennes qui s'accordent avec la *misère d'ici-bas* (*souffrance*, *heureux*, *malheureux*, *abandonné*, *perdu*, *misère*, *misérable*, *poids*, *peser*). Il faut ajouter d'ailleurs que ce problème du bonheur est posé dans les mêmes termes, sinon avec les mêmes conclusions, par Gide dans les trois textes (*Le Voyage d'Urien*, *Les Nourritures terrestres*, *Les Nouvelles Nourritures*) qui figurent sous la rubrique Prose poétique. La tonalité chrétienne est nettement plus accentuée dans la gamme inférieure des fréquences, dont la liste caractéristique se présente comme suit : *hériter*, *ranger*, *éternellement*, *charnel*, *chrétienté*, *pécheur*, *aïeule*, *reniement*, *paroisse*, *périssable*, *veilleur*, *brebis*, *perdition*, *éternité*, *Saint-Père*. Ne dirait-on pas une litanie à l'usage des fidèles de Péguy?

En s'en tenant aux fréquences moyennes et au corpus en vers où l'intervention de Dieu est plus discrète, on peut délimiter les champs sémantiques que féconde le jaillissement poétique. Notons d'abord que le moule est uniforme et que les vocables poétiques sont le plus souvent monosyllabiques. Les 60 premiers de la liste ne contiennent que 7 éléments qui dépassent ce format, il y en a 18, puis 18 et 23 dans les 3 tranches suivantes, tandis qu'inversement les monosyllabes sont l'exception en prose. La poésie s'attache d'abord aux éléments : l'eau (*flot*, *eau*, *mer*, *fleuve*, *neige*, *source*, *goutte*), la terre (*terre*, *sol*, *Pierre*, *sable*, *poussière*, *montagne*, *plaine*, *horizon*, *fer*, *or*), le feu (*feu*, *flamme*, *brûler*, *allumer*, *fumée*), le ciel (*ciel*, *vent*, *nuage*, *pluie*, *soleil*, *lune*, *étoile*). Elle est surtout hypnotisée par les jeux de lumière (*lumière*, *ombre*, *clarté*, *ténèbres*, *lueur*, *rayon*, *voile*, *éclat*, *lampe*, *clair*, *obscur*, *sombre*, *noir*, *blanc*). Elle s'émerveille de la vie des plantes (*forêt*, *arbre*, *bois*, *branche*, *feuille*, *herbe*, *rose*, *fleur*, *fruit*). Et parmi les animaux elle choisit la gent ailée (*oiseau*, *aile*, *plume*). Elle est plus curieuse encore de l'homme, corps et âme. Le corps humain est presque entièrement anobli par la vision poétique : seuls les organes techniques

15. Cette domination de Claudel et Péguy peut correspondre au mérite particulier de ces deux auteurs. Elle est cependant dommageable à l'équilibre statistique du sous-ensemble, les 9 titres des 7 autres auteurs (Gide, Jacob, Saint John Perse, Eluard, Michaux et Ponge) se trouvant écrasés par la masse des 10 textes de Claudel et de Péguy.

comme le nez et l'oreille s'en tiennent à leur fonction utilitaire et prosaïque. Mais le ventre est poétique, comme la chair, le sang et le cœur. Les membres aussi véhiculent l'expression poétique (bras, main, poing, doigt, genou, pied), mais la tête reste le morceau de choix (front, yeux, lèvres, bouche, dent, cou, cheveu). La poésie est donc d'essence charnelle; les verbes — relativement peu nombreux en poésie — ont chance d'être acceptés dans le milieu poétique lorsqu'ils expriment un mouvement du corps, ou un mouvement des corps, comme la pesanteur. L'aspect physique des verbes naître, mourir, dormir, boire, tomber, traîner, danser, rouler, couler, monter, descendre, est ainsi le trait commun qui les rattache à la poésie. Quand l'esprit intervient c'est de préférence dans le contact avec le concret, dans l'acte privilégié de la sensation, qui procure la couleur (vert, bleu, rouge, jaune, blanc, noir), les parfums (odeur, respirer, parfum), ou les sons (silence, bruit, cri, sourd, écouter, taire). À l'adjectif surtout est dévolu le rôle de qualifier le concret poétique et d'en définir la densité, la taille, la forme, l'aspect et le toucher (haut, bas, vaste, large, long, léger, vague, pareil, pur, nu, vierge, doux, dur, frais, froid). L'adjectif est sollicité aussi par la poésie pour qualifier les états d'âme (triste, cruel, calme, noble, tendre, fier, fou, sage, fier, fidèle), avec quelques substantifs et quelques verbes choisis parmi les plus simples (âme, rêve, larme, espoir, douleur, malheur, orgueil, haine, bonheur, joie, paix, pitié, fête, beauté, regret, ennui, douceur, horreur, pleurer, rire, songer, rêver, trembler). Mais au-delà ou au-dedans de l'âme, la poésie s'interdit de pénétrer : ni la métaphysique, ni la psychologie, ni la politique n'ont de permis de séjour en terre poétique. À plus forte raison la science, en tant que reconstruction abstraite du réel, est-elle absente de l'univers poétique, et si le temps apparaît ce n'est pas l'unité de mesure du temps scientifique, mais une tranche particularisée et affective de vie vécue (hiver, printemps, automne, jour, nuit). La poésie est rebelle à la mesure, à la précision et à l'analyse scientifiques et, lorsqu'une mesure quantitative est nécessaire, la poésie utilise l'intensif extrême (infini, immense, profond, ardent, suprême, affreux, énorme), comme ces peuplades primitives qui savent compter jusqu'à deux et qui résument tous les autres nombres dans un pluriel indéterminé. À l'opposé, un regard rapide à la liste de gauche du tableau 2 suffit à désigner les termes prosaïques, dont le caractère abstrait apparaît d'emblée. On peut se demander toutefois si la prose a si mauvais goût qu'elle choisit expressément ces outils encombrants et disgracieux, ces « éponges » lexicales, pour reprendre l'expression de Giraudoux, qui n'aimait guère l'abstraction. En réalité ces éléments sont les laissés pour compte, les rebuts de la poésie. Ne sont significatifs en prose que les mots exclus par la poésie.

Les mêmes conclusions sont suggérées par le tableau III qui reproduit une analyse factorielle des adjectifs les plus fréquents du T.L.F. L'interprétation en est aisée, la contiguïté de deux vocables sur le graphique étant le signe d'une similarité des distributions. Et de la même façon la distance, qui sur le graphique sépare un adjectif d'un genre, est la mesure presque directe de la préférence ou de la répulsion que ce vocable manifeste pour le registre considéré. Enfin la contiguïté ou l'éloignement de deux variables marque la communauté ou l'opposition des genres. Or ces

liens sont constants d'un graphique à l'autre et les nombreuses analyses factorielles que nous avons appliquées au T.L.F. ont la même lisibilité¹⁶. Ainsi voit-on toujours la prose diamétralement opposée à la poésie, de même que Soliloque et Dialogue font équilibre au « Reste » qui se situe toujours dans le quadrant opposé. Prose et poésie diffèrent donc sur les deux facteurs, de même que les personnes. Mais la prose poétique est toujours moins excentrique que les vers et se situe en retrait sur la ligne de tension qui joint les vers et la prose. De même le Soliloque a des exclusives moins prononcées que le Dialogue et se tient moins éloigné de la troisième personne. De plus les deux systèmes d'opposition ne sont pas indépendants puisque la prose et le « reste » partagent toujours, non pas le même quadrant, mais la même moitié, et qu'il en est ainsi du Dialogue et des Vers, et à un moindre degré, du Soliloque et de la Prose poétique.

On peut regretter toutefois que la distinction des genres dans le dictionnaire des fréquences des XIX^e et XX^e siècles n'ait pas séparé le roman et l'essai, la correspondance et les mémoires, le comique et le tragique¹⁷. Une véritable étude statistique des genres reste à faire, sans grande difficulté d'ailleurs, car il suffit pour constituer les effectifs des genres de regrouper les textes appropriés, quitte à délaissier les œuvres dont le genre hybride fait difficulté.

La troisième voie conduit aux auteurs. Ce peut être une synthèse limitée à un seul écrivain, pourvu qu'il soit suffisamment représenté dans le corpus du T.L.F. C'était le cas pour Giraudoux à qui nous avons consacré la première monographie fondée sur les données du T.L.F.¹⁸ C'est généralement le cas aussi des grands auteurs : le T.L.F. incorpore l'œuvre presque complète de Musset, de Stendhal, de Flaubert, de Claudel, de Montherlant, on y trouve intégralement la *Recherche du Temps Perdu*, *Les Hommes de bonne volonté*, *les Thibault*, *Jean-Christophe*, *La Chronique des Pasquier*; Balzac y compte 19 titres, Hugo 19, Lamartine 14, Zola 21, Bernanos 13, Mauriac 13 et Maupassant plus de 200 contes. Rien n'empêche d'ailleurs le chercheur de compléter la gamme et de donner à sa monographie une base plus large, pourvu que les normes du dépouillement et de la lemmatisation soient compatibles avec celles du T.L.F. Il y aurait intérêt également à mettre à l'épreuve les tests statistiques mis au point par Herdan, Guiraud et Muller, en les appliquant à de nouveaux auteurs afin d'apprécier plus sûrement leur validité. Plus nombreuses seront les monographies, plus riches et plus fermes seront les conclusions. Car la sûreté des méthodes statistiques n'est établie que par la convergence des applications et l'originalité d'un écrivain ne peut guère

16. Cf. notre thèse, *Le vocabulaire de Jean Giraudoux, Structure et Evolution*, pp. 528-535.

17. L'exploitation du corpus du XVIII^e siècle s'annonce plus prometteuse sous ce rapport, puisque les textes ont été répartis entre sept rubriques : roman, théâtre, poésie, correspondance, mémoires, voyages et divers.

18. Aux 16 titres retenus par le T.L.F. nous avons cependant ajouté 7 textes dépouillés par nos soins.

être évaluée que par la comparaison avec les autres. En l'absence de monographies déjà constituées, il est cependant possible d'appuyer la comparaison sur un terme de référence, l'ensemble du corpus du T.L.F. ou tel sous ensemble plus proche de l'auteur étudié. Ainsi le vocabulaire caractéristique d'un auteur peut-il être mis en lumière, qu'il s'agisse de ses préférences lexicales ou de ses silences¹⁹. Quel lecteur de Giraudoux n'a pas remarqué l'abondance des adjectifs du bonheur édénique : seul, vrai, premier beau, juste, parfait, pur? Ce sont précisément les éléments que l'ordinateur place en tête de liste.

On peut imaginer une synthèse plus ambitieuse qui réunirait dans la même étude un certain nombre d'écrivains en choisissant de préférence ceux qui appartiennent à la même époque et au même genre afin d'éliminer les phénomènes perturbateurs. Ainsi isolée l'originalité propre à chaque auteur devrait apparaître par contraste. Mais on a le soupçon qu'une telle étude pourrait ne pas se borner à souligner la diversité des tempéraments littéraires, et devrait mettre en évidence ce qui rapproche les écrivains : non pas seulement les contraintes de la langue et du genre, ou la dominance des idées ou des thèmes liés à tel milieu, à tel pays, à telle époque, mais aussi ce qui semble bien être la loi de la nature, la marque du temps et de l'âge. La succession chronologique des textes ne fait-elle pas apparaître des constantes qu'on retrouverait d'un écrivain à l'autre? Quelle est l'influence des années sur le langage et sur la création? Est-ce qu'en vieillissant un auteur s'appauvrit? Gagne-t-il en profondeur et en sobriété ce qu'il perd en surface et en dispersion? Les mots qu'il utilise ne gagnent-ils pas en extension ce qu'ils perdent en compréhension? L'étude de Giraudoux nous a amené à nous poser ces questions. Rien ne laissait présager la marque du temps chez cet auteur svelte et souriant qui, comme Cocteau, faisait profession de jeunesse. Et pourtant les chiffres montrent qu'une évolution irrésistible le conduit de la richesse à la sobriété, de la fantaisie à la gravité, du concret à l'abstrait, de la poésie au réel, de la nature à l'homme, du nom au verbe. On est bien obligé d'admettre que le corps change au cours des ans. Pourquoi l'esprit échapperait-il au sort commun? On sait quel vertige Proust éprouvait devant le temps qui coule et avec quelle obstination il tentait de jeter son ancre dans le courant. Sur quelle distance le flot l'a-t-il malgré lui emporté? L'arsenal statistique appliqué à son œuvre devrait pouvoir éclairer ce problème aussi bien qu'une batterie d'encéphalogrammes.

Ce sont là quelques directions de recherche comparative offertes à qui veut traiter en bloc ou en partie les données²⁰ du T.L.F. L'exploitation de la carrière est à peine commencée, et le filon ne s'épuisera pas de sitôt. Loin d'en être diminué, le Trésor en tirera un intérêt supplémentaire, même si la banque de données prête à fonds perdu.

19. La méthode est simple : quand on connaît la fréquence d'un mot dans le corpus et chez l'auteur étudié, une simple règle de trois permet le calcul de la fréquence théorique. 11 suffit alors de mesurer l'écart et de le réduire à l'échelle commune.

20. Les données du T.L.F. méritent bien leur nom. Elles sont communiquées à titre gracieux ou pour un prix modique.

TABLEAU II. LE VOCABULAIRE SIGNIFICATIF DES GENRES
DANS LE T.L.F. (XX^e siècle)
par ordre décroissant de signification (fréquences moyennes)

PROSE	écart réduit	PROSE POÉTIQUE	écart réduit	VERS	écart réduit
GÉNÉRAL	18.33	TERRE	38.51	CIEL	68.07
ÉTAT	16.35	ÉTERNEL	36.97	OMBRE	58.50
FRANÇAIS	14.99	EAU	34.51	AME	46.88
EFFET	14.65	SANG	33.04	VENT	46.83
POLITIQUE	14.34	SAINTE	32.99	OR	44.66
LETTRE	14.12	OISEAU	31.75	ROSE	42.66
SENTIMENT	14.10	ESPÉRANCE	29.72	FLEUR	41.69
CAS	14.06	CIEL	28.59	SOLEIL	39.64
QUESTION	13.69	MER	25.26	AILE	35.44
COMPRENDRE	13.56	CONNU	23.88	SOMBRE	35.42
SUITE	13.41	SOLEIL	23.58	EAU	34.68
ÉCRIRE	13.19	PEUPLE	21.45	DIEU	34.09
PARTIE	12.88	PAIN	19.35	SANG	33.36
CONDITION	12.86	FRUIT	18.80	LUNE	33.29
AGIR	12.85	DORMIR	18.55	FRONT	33.14
RÉPONDRE	12.66	SOURCE	18.33	DOUX	33.06
ACTION	12.59	PÉCHÉ	18.27	FLOT	32.46
ALLEMAND	12.55	PIED	18.15	HÉLAS	31.72
POSSIBLE	12.43	FILS	17.94	NOIR	31.61
MANIÈRE	12.28	SABLE	17.71	ÉTOILE	31.52
PERSONNE	12.22	AME	17.26	OISEAU	30.22
CONSCIENCE	12.21	FLEUR	17.17	MER	30.17
SOCIÉTÉ	12.16	POÈME	15.74	ANGE	29.81
PARAITRE	12.08	ANTIQUÉ	15.35	CHAIR	29.36
EXPÉRIENCE	11.85	ROSE	14.73	FRUIT	27.92
CONTRAIRE	11.84	PRIÈRE	14.72	CHANTER	27.35
AFFAIRE	11.82	MAIGRE	14.35	NU	26.62
PIÈCE	11.76	IMMENSE	14.30	VIERGE	26.32
MINUTE	11.74	LUNE	13.87	FLAMME	26.25
ABORD	11.72	CROIX	13.74	VIN	25.99
ARMÉE	11.69	FEUILLE	13.71	MOURIR	25.76
SUJET	11.68	PLEURER	13.58	TERRE	25.25
DOUTE	11.56	ARBRE	14.43	ÉTERNEL	25.11
ORDRE	11.49	SAGESSE	13.22	VAIN	24.79
RAPPORT	11.39	CRÉATION	13.01	BAISER	24.76
FRANC	11.33	NEIGE	12.82	RÊVE	24.65
EXPLIQUER	11.29	PIERRE	12.73	ROI	24.56
SOCIAL	11.22	PLEIN	12.59	FEU	24.53
JOURNAL	11.19	CHANTER	12.50	PALE	24.49
CARACTÈRE	10.97	SALUER	12.45	PLAINÉ	23.87
		ANGE	12.43	CLARTÉ	23.35

PROSE	écart réduit	PROSE POÉTIQUE	écart réduit	VERS	écart réduit
INTÉRÊT	10.95	SEIGNEUR	12.22	TRISTE	22.40
VALEUR	10.87	ROI	12.21	PUR	21.73
MORAL	10.57	PLUIE	12.11	HORIZON	21.53
VUE	10.57	JARDIN	12.10	CLAIR	21.44
SITUATION	10.43	CHEMIN	11.68	PLEURER	21.33
SORTE	10.38	CRÉATURE	11.63	TREMBLER	20.53
OBJET	10.38	PUR	11.39	SEIGNEUR	20.36
FAÇON	10.31	FLEUVE	11.31	LOURD	20.12
PROBLÈME	10.09	CHANT	11.26	SEIN	20.01
ARTICLE	10.08	ÉTOILE	11.07	GLOIRE	20.00
ENTENDU	10.01	DOUBLE	11.01	SOUFFLE	19.81
EPROUVER	10.00	SEUIL	10.98	FEUILLE	19.72
CONVERSATION	10.00	VIN	10.96	FLEUVE	19.54
		RAYON	10.78	FUIR	19.42
IMPRESSION	9.99	CORPS	10.69	DIVIN	19.39
DINER	9.99	FLOT	10.60	DORMIR	19.24
NATUREL	9.99	POUSSIÈRE	10.58	LONG	19.13
ACTE	9.91	SOUFFRANCE		CRI	19.09
SALON	9.70		10.54	PLEIN	19.04
PARTI	9.59	NU	10.51	BLANC	19.01
RELATION	9.52	FACE	10.45	CROIX	18.85
EXPRESSION	9.52	RACE	10.28	DOIGT	18.80
MOUVEMENT	9.50	BLANC	10.10	BOIS	18.76
CRITIQUE	9.49	CHAIR	9.93	VILLE	18.72
PENSÉE	9.47	DESERT	9.87	FORET	18.63
MOYEN	9.45	ORGUEIL	9.86	FOND	18.63
EXEMPLE	9.45	OR	9.80	RAYON	18.43
EGARD	9.43	BORD	9.76	INFINI	18.35
CONNAISSANCE	9.43	ENORME	9.71	VOILE	17.99
		CHRETIEN	9.60	BOIRE	17.81
EXISTENCE	9.41	HEUREUX	9.52	BOUCHE	17.60
SERVICE	9.39	VAIN	9.30	HIVER	17.51
SENS	9.38	HERBE	9.23	ARDENT	17.34
OPINION	9.36	PRINTEMPS	9.20	PRINTEMPS	17.27
VISITE	9.36	MALHEUREUX		REINE	17.27
ANGLAIS	9.36		9.17	ADIEU	16.95
PARTICULIER	9.35	COULEUR	9.10	CHANT	16.89
BESOIN	9.35	VIVANT	9.09	SONNER	16.28
CONTINUER	9.33	VALOIR	9.03	DESTIN	16.20
PHILOSOPHIE	9.31	FAIM	8.90	ANTIQUE	16.13
HISOIRE	9.29	LUMIERE	8.89	etc...	
		CHERCHER	8.87		
		UNIQUE	8.85		
		etc...			

Une liste plus complète figure en annexe à notre thèse : pp. 1107-1132.

TABLEAU III.

Analyse factorielle des adjectifs du T.L.F.

