

recourant, pour prendre leur revanche, à la violence masculine des 'cognati,' du 'buono uomo.' De la même manière, la 'bouche absente' des textes semble devoir toujours emprunter, pour parler, la main écrivante d'un autre, quelle que soit par ailleurs la fascination directe qu'ils exercent.¹⁶

University of Michigan

Novalis: Une autre Vision du texte

Je me propose de vous présenter à grands traits les divers moments de notre projet¹ où nous avons eu recours à l'assistance de l'ordinateur.

Dans un premier temps, je vous exposerai les raisons intrinsèques à notre objet et à notre méthode qui justifiaient le recours à une approche informatisée, ensuite j'aborderai l'aspect pratique de la question et les choix que nous avons été amenés à faire et je conclurai en évoquant quelques implications théorico-méthodologiques du traitement automatique du texte.

Le traitement automatique des textes était à ses débuts pratiqué dans des cercles restreints; progressivement cependant l'attitude du 'monde littéraire' à son égard est devenue moins hostile ou réticente. Depuis ces dernières années, l'accueil est franchement positif et intéressé vu les énormes avantages qu'offre cette approche du texte. Il est révolu le temps où seuls les informaticiens et les statisticiens s'adonnaient au traitement automatique du texte, étant bien incapables d'imaginer comme 'critique littéraire' autre chose que d'interminables séries de formules nébuleuses, de calculs d'indices hermétiques et de listes de chiffres et de mots qu'ils étaient en peine d'interpréter. De nos jours, les meilleurs centres de traitement de texte fonctionnent autour d'une équipe multidisciplinaire formée d'informaticiens et de littéraires. La production automatique de diverses listes de vocabulaire et de statistiques qui était une fin en soi est devenue maintenant un moyen pour épauler des recherches précises; néanmoins, cette démarche exige tant de l'informaticien que du littéraire une bonne connaissance de l'autre discipline.

* * *

¹⁶ C'est le moment de dire certaines des voix auxquelles se substitue ici la mienne. Les principaux thèmes de l'essai qu'on vient de lire sortent des discussions de deux séminaires (University of Michigan, automne 1983 et hiver 1984). Décisive fut l'intervention d'Ian Reid, qui concurremment attirait mon attention sur l'aventure de Fra Alberto et me faisait réfléchir, à travers des analyses flaubertiennes, au caractère sacrificiel du récit. Peut-être n'ai-je fait que gloser sa formule: 'No story is the whole story.' (Voir, à ce propos, la préface de l'ouvrage collectif qu'il a édité avec Sneja Gunew: *Not the Whole Story* [Sydney: Local Consumption Publications 1984].)

¹ Projet de recherche subventionné par le Conseil de Recherches en Sciences Humaines du Canada, dirigé par Prof. Walter Moser et intitulé 'L'interaction des discours chez Novalis et Musil.' Le présent article est inspiré largement par le texte de la communication que j'ai donnée aux Sociétés Savantes à Guelph en juin 1984 lors de la séance 'Comparative Literature and the New Paradigms,' conjointement avec Walter Moser et Denise Ringnette au cours d'une triple intervention intitulée 'L'interdiscursivité du texte littéraire: Novalis et Musil.'

Pourquoi le traitement automatique du texte? On peut douter en effet de l'opportunité du traitement automatique du texte en ce qui concerne bon nombre de recherches portant sur des textes courts. Par contre, pour des corpus de textes plus nombreux, plus longs ou plus compliqués l'approche informatisée devient un détour intéressant, voire indispensable au cours du cheminement interprétatif.

Le texte du Brouillon de l'Encyclopédie de Novalis, qui est un cas limite, et le type de recherche que nous poursuivons illustrent de façon fortuite mais idéale le pourquoi de notre détour informatique, essentiellement pour trois raisons corollaires:

1) La taille du texte est relativement conséquente à savoir: un peu plus de 270 pages de 23 lignes en moyenne, pour un total de 58 578 mots ou 540 000 caractères typographiques. Beaucoup parmi ces mots ont une longueur très germanique pouvant atteindre même 40 lettres² (L5772-ZWEITE) ou 42 caractères. Voir exemple 1.

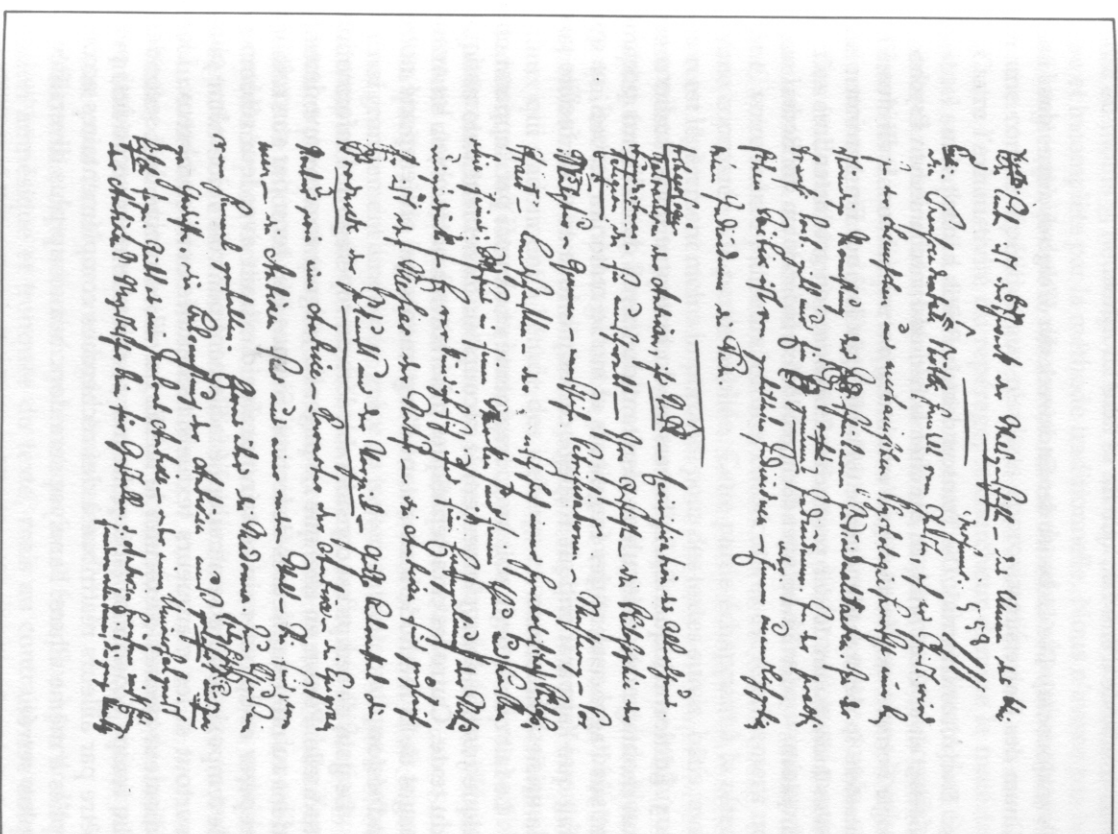
EXEMPLE 1

(Ihre Wissenschaft) ist eine Algeber der Physik und Technologie.) < iZeichenflu = *chenform(figuren)bedeutungskunst* > i. (Die jetzt sog[enannte] Algeber ist schon ein U = bergang zur speziellen Arithmetik — weil Zahlen dabey vorkommen.) Acustische < iPerspectivformen > i oder Figuren.

2) Texte fragmenté: Même si beaucoup d'entre nous ont déjà travaillé sur des ouvrages de taille équivalente, le texte de Novalis est constitué de 1151 fragments qui ne présentent entre eux que peu ou pas de liens apparents au niveau de l'enchaînement syntagmatique. Il en découle que le texte est difficilement mémorisable et peu commode pour le repérage d'information. Voir exemple 2.

² La différence entre le nombre de caractères et de lettres est imputable au fait que les lettres ambiguës ou spéciales sont codées par des combinaisons de plusieurs caractères, par exemple le ß est codé < s ou le ä par a=.

EXEMPLE 2



Faksimile der Handschrift R, Blatt 3b
(verkleinert; vgl. S. 248, Z.4-27)

Im Besitz des Freien Deutschen Hochstifts, Frankfurt a.M.

Chemische und mechanische Psychologie. Transscendentale Poetik. practische Poetik. Die Natur zeugt, der Geist macht. Il est beaucoup plus commode d'être fait, que de se faire lui même.

PSYCH[OLOGIE]. Die Liebe ist der Endzweck der *Weltgeschichte* – das Unum des Universums.

5

51. ENC[YCLOPAEDISTIK]. Die Transcendentale Poetik handelt vom Geiste, eh er Geist wird. In der chemischen und mechanischen Psychologie herrscht eine beständige Vernichtung der scheinbaren Individualitäten. In der transcendentalen Poetik gibt [es] nur Ein *gemeines rohes Individuum*. In der practischen Poetik ist von gebildeten Individuen – oder Einem unendlich gebildeten Individuum die Rede.

10

3) Édition critique: Enfin la version du texte que nous étudions est un montage très méticuleux résultant du travail des éditeurs portant et sur l'agencement des fragments et sur le matériau textuel, ce qui fait que l'apparat critique n'allège certes pas le texte et ne facilite pas la saisie à la lecture.

Le fait que le texte soit long, fragmenté et alourdi par l'apparat critique justifie donc pleinement le recours au traitement automatique du texte. Outre les caractéristiques inhérentes à notre objet, les avantages de la méthode sont autant d'arguments qui ont emporté notre adhésion.

Le gain de temps est considérable lors du repérage de l'information textuelle. Pensez en effet que 270 pages de 35 lignes en moyenne demandent au moins 25 heures de lecture. Si vous êtes tenu par surcroît, de repérer et de noter un certain nombre d'extraits avec leur référence, le temps de lecture sera multiplié facilement par trois voire même plus, surtout si les marqueurs textuels recherchés sont nombreux. L'ordinateur permet d'exécuter le même travail en moins de 5 secondes. Le temps économisé (soit 74 heures 59 minutes et 55 secondes) peut être par ailleurs réattribué à des recherches complémentaires accomplies à même allure: l'analyse sera donc beaucoup plus diversifiée et plus serrée.

Au gain de temps correspond un gain d'argent car il faut avoir une évaluation bien inconsciente du coût horaire d'un chercheur d'institution universitaire, surtout en ces temps de restriction, pour s'entêter à recourir à une méthode traditionnelle coûteuse et lente.

Le repérage automatique de toutes les occurrences d'un mot ou d'une

configuration de mots dans tel ou tel contexte ou à l'exclusion de tel autre contexte prend toujours moins de cinq secondes alors qu'il aurait fallu au moins 75 heures pour faire le même travail de manière imprécise et incomplète par la méthode traditionnelle. Nous n'avons pas hésité longtemps, je dois le dire, entre une recherche exhaustive et rapide et une corvée répétitive aux résultats approximatifs.

Outre l'exhaustivité du repérage, le chercheur domine le matériau textuel sur lequel il a maintenant une emprise totale et directe. Les éléments du texte sont maléables à volonté, repérables et susceptibles d'être triés dans n'importe quel ordre; aucun ne s'égare dans l'ombre de notre mémoire défaillante: tout est désormais à notre portée.

Des études récentes à propos de la rétention d'information ont montré que la fraction du texte dont le lecteur est capable de se souvenir après une, voire même plusieurs lectures consécutives est relativement minime comparée à la partie oubliée. Cette partie échappant à la rétention est légèrement moins importante pour des textes suivis, bâtis, avec une trame narrative que pour des textes lyriques de même longueur. À plus forte raison, que dire alors du texte hyper-fragmenté de Novalis qui est tel un iceberg dont la partie immergée dans les oubliettes de notre mémoire est infiniment plus grande que la partie émergée. Ceux qui s'adonnent à l'étude des icebergs en ne considérant que la partie hors de l'eau, manquent, à mon sens, de rigueur. Il me semble que le traitement automatique du texte permet précisément d'éviter d'avoir une vision partiellement amnésique du texte.

La qualité de la perception que nous avons de notre objet se trouve aussi grandement améliorée: grâce au traitement automatique permettant un retour facile et immédiat au texte pour fin de vérifications, il devient presque impossible de sous-évaluer ou de sur-évaluer l'ampleur de tel ou tel phénomène littéraire. Nous disposons donc d'un moyen sûr pour éviter d'avoir du texte une vision tronquée, capables que nous sommes maintenant d'assurer le lien ou l'ancrage au texte pour toute hypothèse interprétative.

Enfin, dépendamment de la grille de lecture ou de la méthode choisie, le matériau textuel était tamisé systématiquement. Seul ce qui était pertinent pour l'analyse envisagée était retenu.

Désormais, cette sélection systématique ne se fera plus à partir d'une vision amnésique et tronquée du texte, mais au contraire en tenant compte de tous les éléments du texte. En conséquence, on se doit maintenant de justifier d'une manière ou d'une autre la non-sélection des éléments du texte qui ne sont pas retenus par l'analyse et de dire le statut particulier qu'ils occupent au sein du matériau textuel.

* * *

À ce stade de réflexion lors de la préparation du projet, il s'agissait pour nous de prendre une série de décisions pratiques:

1) Le corpus fragmenté appelé par les éditeurs 'Le Brouillon Général: matériaux pour une encyclopédistique' a connu plusieurs éditions différentes. Le critère de sélection a été moins l'ordre des fragments que l'aspect 'complétude' du corpus. Notre choix s'est donc porté sur le travail de Hans Joachim Mähl³ qui n'est pas une sélection parmi un réservoir de fragments textuels mais bien un effort pour présenter l'ensemble de manière chronologique.

2) Un texte, entré en mémoire tel quel, est une suite d'impulsions stockées sur un support magnétique. Le codage introduit un peu d'ordre au sein de ces informations. Voir exemple 3. Certains codes indiqueront les caractéristiques physiques du texte: marques de page, de ligne, de fragment et de partie pour faciliter la référence à l'immediate.⁴ S'ajoutent alors des codes morpho-syntaxiques: marques de phrase et de ponctuation et des codes orthographiques.⁵ Enfin, toutes les traces du travail d'édition sont aussi reprises: marques des différentes fontes de caractères,⁶ des signes d'édition, d'abréviation,⁷ de souligné, etc.

Ainsi tous ces codages concourent à nous fournir une copie magnétique du texte aussi riche en informations que le manuscrit. Voir exemple 3 et exemple 4.

3) Travaillant avec le laboratoire de traitement de texte de Serge Lusinjan et Paul Bratley, nous avons aussi bénéficié de l'assistance de leur informaticien spécialisé, Bernard Derval. Les logiciels dont ils disposaient pouvaient facilement traiter des textes allemands de la taille du Brouillon. Nous avons donc pu avoir accès à des programmes de validation des données après l'enregistrement, à des programmes de

3 'Das Allgemeine Brouillon, Materialien Zur Enzyklopädistik 1798/1799,' in *Novallis Schriften*, Dritter Band. Das Philosophische Werk II. Herausgegeben von Richard Samuel in Zusammenarbeit mit H. J. Mähl und G. Schulz, (Stuttgart: Kohlhammer Verlag 1975) 242-478

4 1.246.01 correspond à: première partie, page 246, ligne 1.

5 À la ligne 1.246.08 nous avons *la* < st dont nous venons d'expliquer le codage.

6 L'italique est toujours précédé du code < i afin de toujours pouvoir faire des tris de vocabulaire ou des statistiques uniquement sur les mots indiqués en italiques dans le texte.

7 Les crochets [] indiquent les parties que les éditeurs ont ajoutées au texte original de Novalis. Les chercheurs qui désirent analyser le travail d'édition de Mähl peuvent grâce à ces codes facilement lister tous les mots ou toutes les phrases complétées par les éditeurs.

EXEMPLE 3

1640=1.246.01 49. PSYCH[OLOGIE] UND ENCYKLOP[AEDISTIK]. <iDeutlich>i wird etwas
1650=1.246.02 nu[r] [du]rch Repraesentation. Man versteht eine Sache am Leicht-
1660=1.246.03 t[este]n, wenn man sie repraesentirt sieht. So versteht man das Ich
1670=1.246.04 nur insofern es vom N[icht]I[ch] repraesentirt wird. Das N[icht]I[ch]
1680=1.246.05 ist das Symbol des Ich, und dient nur zum Selbstversta=ndni<s des Ich.
1690=1.246.06 So versteht man das N[icht]I[ch] umgekehrt, nur insofern es vom Ich
1700=1.246.07 repraesentirt wird, und dieses sein Symbol wird. In Hinsicht auf die
1710=1.246.08 Mathem[atik] la= < st sich diese Bemerkung so anwenden, da<s die
1720=1.246.09 Mathem[atik], um versta=ndlich zu seyn repraesentirt werden mu<s.
1730=1.246.10 Eine Wissenschaft la= < st sich nur durch eine andre wahrhaft repraesen-
1740=1.246.11 tiren. Die paedagogischen Anfangsgru=nde der Mathem[atik] mu=ssen
1750=1.246.12 daher < isymbolisch > i und < ianalogisch > i seyn. Eine bekannte W[issenschaft]
1760=1.246.13 mu<s zum < iGleichni<s > i fu=r die Mathematik dienen und diese Grundglei-
1770=1.246.14 chung mu<s das Princip der Darstellung der Mathematik werden./
1780=1.246.15 So wie die Anthropologie, die Basis der Menschengeschichte, so ist die
1790=1.246.16 Physik der Mathematik die Basis der Geschichte der Mathematik.
1800=1.246.17 Die Physik u=berhaupt ist die urspru=ngliche, eigentliche Geschichte.
1810=1.246.18 Die gewo=hnlich so genannte Geschichte ist nur [abg]eleitete Ge-
1820=1.246.19 schichte./
1830=1.246.20 / Gott selbst ist nur durch [Re]praesentation versta=ndlich./

EXEMPLE 4

49. PSYCH[OLOGIE] UND ENCYCLOP[AEDISTIK]. *Deutlich* wird etwas
 nu[r] [du]rch Representation. Man versteht eine Sache am leicht-
 [teste]n, wenn man sie representirt sieht. So versteht man das Ich
 nur insofern es vom N[ic]ht[ic]h] representirt wird. Das N[ic]ht[ic]h]
 ist das Symbol des Ich, und dient nur zum Selbstverständnis des Ich.
 5 So versteht man das N[ic]ht[ic]h] umgekehrt, nur insofern es vom Ich
 representirt wird, und dieses sein Symbol wird. In Hinsicht auf die
 Mathem[atik] läßt sich diese Bemerkung so anwenden, daß die
 Mathem[atik], um verständlich zu seyn representirt werden muß.
 Eine Wissenschaft läßt sich nur durch eine andre wahrhaft represen-
 10 tieren. Die paedagogischen Anfangsgründe der Mathem[atik] müssen
 daher *symbolisch* und *analogisch* seyn. Eine bekannte W[issens]schaft
 muß zum *Gleichniß* für die Mathematik dienen und diese Grundglei-
 chung muß das Princip der Darstellung der Mathematik werden./
 15 So wie die Anthropologie, die Basis der Menschengeschichte, so ist die
 Physik der Mathematik die Basis der Geschichte der Mathematik.
 Die Physik überhaupt ist die ursprüngliche, eigenliche Geschichte.
 Die gewöhnlich so genannte Geschichte ist nur [ab]geleitete Ge-
 schichte./
 20 / Got selbst ist nur durch [Re]presentation verständlich./

49. PSYCHOLOGIE ET ENCYCLOPÉDISTIQUE

Une chose ne devient *claire* qu'au moyen de la représentation. On comprend
 une chose plus facilement quand on la voit représentée./ Aussi, ne comprend-
 on le Moi que dans la mesure où il est représenté par le Non-Moi. Le Non-
 Moi est le symbole du Moi et ne sert qu'à la compréhension du Moi. Aussi,
 inversement, ne comprend-on le Non-Moi que dans la mesure où il est
 représenté par le Moi et en devient le symbole./ Cette réflexion peut s'appli-
 quer aux mathématiques si bien que pour être intelligibles, elles doivent être
 représentées. Une science ne peut être vraiment représentée que par l'inter-
 5 médiaire d'une autre./ Les premières notions de pédagogie des mathématiques
 doivent, par conséquent, être *symboliques* et *analogiques*. Une science connue
 doit servir de *figure analogique* aux mathématiques et cette équation fonda-
 mentale doit devenir le principe d'exposition des mathématiques./ De même que
 l'anthropologie est le fondement de l'histoire du genre humain, la physique
 des mathématiques est le fondement de l'histoire des mathématiques. La phy-
 10 sique, quoi qu'il en soit, est l'histoire véritable et originelle. Ce qu'on appelle
 communément l'histoire n'en est qu'un dérivé./
 /Dieu lui-même n'est intelligible que grâce à la représentation./

préparation du texte, à des programmes de repérage et de tri des in-
 formations et enfin des programmes de comptage et de statistiques.

Notre démarche de recherche est simple: à la suite de la mise au point
 du cadre méthodologique et théorique, nous avons retenu un certain
 nombre d'hypothèses interprétatives et déterminé une série de con-
 cepts opérationnels. Ensuite, nous avons choisi des marqueurs textuels
 qui devraient illustrer notre hypothèse à même le matériau du texte.
 Enfin, après avoir procédé au repérage des marqueurs choisis de
 manière automatique, nous analysons les listes de résultats afin de re-
 10 jeter ou corroborer nos hypothèses.

Le processus informatisé comprend les trois étapes suivantes:

1) L'entrée-validation des données est sans doute la partie la plus
 fastidieuse du travail. L'originalité technologique de notre projet est
 d'avoir eu recours au lecteur optique qui nous a épargné la tâche de
 devoir dactylographier notre texte sur un terminal.

2) Une fois le 'magnétoécrit' corrigé, nous produisons diverses listes
 qui seront autant d'outils de travail efficaces.

D'abord une liste des mots du texte triés par fréquence. Voir exem-
 15 ple 5. Cette liste présente dans la colonne de gauche une série de chiffres
 qui sont en fait la première adresse du mot dans le texte. Si l'on sélec-
 tionne le mot *Symbol*, il se trouve la première fois dans la première partie
 du Brouillon, à la page 246 et à la ligne 5(1.246.05). La colonne de chiffre
 à la droite de la liste correspond à la fréquence absolue du mot *Symbol*,
 en d'autres termes, le nombre de fois — dans ce cas-ci 13 fois — qu'il
 apparaît dans le texte sous cette forme précise. Grâce à l'adresse du
 mot, je puis directement retourner au texte pour étudier le passage
 (voir exemple 3) où se trouve le mot.

Dans le domaine des études de fréquences, il y a deux écoles de pen-
 sée qui s'affrontent: l'une se concentre sur l'examen des mots de fortes
 fréquences qui, dans cette optique, sont les plus importants du texte
 et l'autre ne regarde que les mots de basses fréquences, dont la rareté
 semble être la garantie de leur importance. Souvent aussi, le calcul des
 fréquences se fait par rapport à la fréquence du mot dans la langue
 en général: les statisticiens produisent dans ce cas des listes d'écarts
 par rapport à la norme (dictionnaire des fréquences); ainsi le chercheur
 peut repérer les mots qui ont, dans le texte étudié, une fréquence d'utili-
 20 sation très différente de celle qu'ils ont dans le corpus de la langue
 en général.

Ensuite, nous produisons un dictionnaire de tous les mots du texte,
 non-lemmatisés. Voir exemple 6. La présentation du dictionnaire est
 un peu différente de la liste précédente; en effet, la colonne de gauche
 présente la liste des mots du texte triés alphabétiquement, tel un dic-

tionnaire. De plus, les rangées de chiffres disposés sur 6 colonnes constituent le répertoire de toutes les adresses de chacun des mots du texte; le système de référence est le même que pour les listes de fréquences (partie, page, ligne). La colonne complètement à droite de la liste indique la fréquence du mot. La différence entre cette liste-ci et la précédente est qu'elle spécifie toutes les adresses d'un mot et de plus la clef de tri n'est plus ici la fréquence mais bien l'ordre alphabétique des mots. Ceci présente l'immense avantage de pouvoir, d'un seul coup d'oeil, évaluer l'importance de toute la famille sémantique d'un mot. Reprenons l'exemple du mot *Symbol*: il est accompagné dans la même colonne du dictionnaire par toutes les variantes de genre, de nombre, de catégorie grammaticale ou de personne que la racine *Symbol* peut présenter dans le texte. Auparavant, chacune de ces variantes – allant de *Symbol* à *Symbols* – était dispersée dans les listes de fréquences puisque chacune d'elles n'a pas obligatoirement la même fréquence (13 occurrences pour *Symbol*). Les listes d'adresses permettent de nouveau un retour facile au texte afin de sélectionner les mots ou les phrases intéressantes.

Enfin, nous établissons une concordance courte, c'est-à-dire avec un contexte limité une ligne de texte, pour tous les mots du Brouillon. Voir exemple 7. La colonne de gauche indique l'adresse du mot et de sa phrase de contexte. La référence permet également le retour au texte imprimé et pas uniquement à la liste du texte enregistré sur un support informatique puisque l'on respecte toujours lors de l'entrée du texte le même découpage de page et de ligne pour les deux états du texte (livre et magnéto). La colonne centrale de la concordance met en vedette les mots recherchés de la famille du mot *Symbol* et de part et d'autre du mot vedette se répartit le contexte immédiat, soit au maximum 80 caractères.

Ces outils sont des voies d'exploration du texte; suivons, si vous le voulez, notre manière de procéder au cours de cette étape. Lors de l'étude du texte (exemple 3), nous repérons des termes importants pour une investigation précise, c.-à-d. des mots, des éléments du texte qui illustrent un concept d'analyse, qui incarnent un thème à même le matériau textuel. Prenons par exemple le mot '*Symbol*'. Nous le retrouvons dans l'exemple 5 dans la liste des mots de vocabulaire triés par fréquence et dans l'exemple 6 dans le dictionnaire où il se retrouve avec toute la famille de mots de même racine. En suite, un rapide coup d'oeil dans la concordance KWIC nous donne une petite idée du genre de contexte entourant le mot '*Symbol*', comme montré à l'exemple 7. Si le filon s'avère fructueux, nous retournons au texte grâce aux référer-

EXEMPLE 7

NOVALIS

01 juin 1984

20:58

1

1.246.05 wird. Das N[icht][ch] ist das
1.246.07 praesentirt wird, und dieses sein
1.263.08 elphi>i. / PHYS[IK]. Licht
1.271.24 – Sie werden <ioeliger>i. (Oel,
1.321.14 herrlich mit dem Paradiese, dem
2.397.01 [685.] U=ber den Anzug – als <i
2.397.24 ngen der Alten etc. Kleidung <=
2.397.26 heil der Tropik ausmacht. Jedes
2.398.06 Pflanzewelt] – (Alles kann
2.399.27 logie>i. 690. Die Frau ist das
2.399.28 <iScho=nheit>i – der Mann das
3.434.05 mation und Corporation zugleich.
4.477.01 \ [1145.] D[ie] Blu=the ist das
2.397.27 gegensymbole. Es giebt aber auch
2.397.28 Es giebt aber auch Symbole der
3.434.03 ion ist eine Tropisation – oder
3.441.05 und Bilderspiel – Begleitende
1.246.12 der Mathem[atik] mu=ssen daher <i
1.272.32 – eine <ikritische Bemerkung>i (
1.272.32 itische Bemerkung>i (Symbolisch
1.273.21 d=werdung – / ARTISTIK.
1.321.17 gleichsam die gro=se Geschichte
1.248.16 haften – der [Me]nshat immer
1.263.22 hin=deutung auf <iAstrologie>i.
2.398.06 kann Symbol des Andern seyn
4.457.26 chspiel a=hnlichen <iSpiels>i –
1.252.05 stem der Wissenschaften>i soll <i
2.399.35 u=te Scho=n=heit nothw[endig] <i
3.403.08 plettirt und umg[ekehrt] . Beyde
2.397.27 durch sein Symbolisirtes wieder
2.397.29 chselung des <iSymbols>i mit dem
2.397.26 t. Jedes Symbol kann durch sein
2.397.25 s Vorhergehende geho=rt in die <i
2.398.04 <iPlaneten>i und Sternbilder.)
4.450.02 i= <imenologie – Grammatik>i –
2.397.29 ymbole. Auf Verwechselung des <i

Symbol
Symbol
Symbol
Symbol
Symbol
Symbol
Symbol
Symbol
Symbol
Symbol
Symbol
Symbole
Symbole
Symbolisation
symbolisch
Symbolisch
symbolisch
symbolisch
symbolische
Symbolische
symbolische
symbolischer
symbolisiren
Symbolisiren
Symbolisirt
Symbolisirten
Symbolisirtes
Symbolistik
Symbolistik
Symbols

des Ich, und dient nur zum Selb
wird. In Hinsicht auf die Math
und Agens der <iReinheit>i. Wo
des Geistes – sein Ko=rper.)
der Jugend an und schlie<st mit
. Die Scherpe d[es] Kindes ist
des Geistes d[er] Zeiten. Das
kann durch sein Symbolisirtes wi
des Andern seyn – Symbolische
der <iGu=te>i und <iScho=nheit>
der Wahrheit und des Rechts. W
ist individuelles, <itropisches>i
des Geheimnisses unsers Geistes
der Symbole – Untersymbole. A
– Untersymbole. Auf Verwechse
zugleich – eine a=chte Con=st
oder <iSchematism>i. Emanat[io
und <ianalogisch>i seyn. Eine be
symbolisch nach der Lehre v[on]
nach der Lehre v[on] den Signat
religiöse Mimick – Sittenmimi
verju=ngt liegen.) Der Anfang d
Philosophie seines Wesens in se
Prophezeungen. Chiromantie. 1
Function.) <ider Natur>i – der
Gedankenkonstruktionen zu stande
Ko=rper>i (Organsystem) unsers
und <isignalisiren>i zugleich. \
sich einander. Wer weis, was phi
werden – Gegensymbole. Es giebt
– auf ihre Identisirung – auf
wieder Symbolisirt werden – Ge
– die einen Theil der Tropik a
d[es] <imens[lichen] Ko=rpers>
. / Wir sehn d[ie] Natur, so wie
mit dem Symbolisirten – auf ihr

ences se trouvant dans chaque liste pour faire une étude plus approfondie des passages (voir exemple 4) où se trouve le mot recherché.

3) Lors du travail d'exploitation et d'analyse à proprement parler, nous consultons systématiquement ces listes-outils et si la voie est intéressante, nous produisons des concordances avec un contexte beaucoup plus long, pour plusieurs mots d'un même champ sémantique ou bien en rapport avec un même point d'analyse. Ainsi pouvons-nous étendre notre prise sur le texte. En effet, nous avons la possibilité de produire des concordances longues, où le contexte entourant le mot peut compter de quatre à cinq lignes de texte, de telle sorte que lors d'une recherche précise, nous ayons à portée de la main tous les extraits du Brouillon dans lesquels figurent les mots recherchés. Ainsi peut-on par exemple étudier l'évolution d'un thème ou d'un concept au travers de l'œuvre.

Au vu de ces diverses listes et après étude des extraits, les membres de l'équipe de recherche sont à même de vérifier ou d'invalider les hypothèses interprétatives qu'ils avaient retenues au départ. De plus, nous avons à notre disposition un outil de premier ordre pour vérifier les interprétations d'autres chercheurs en essayant de voir jusqu'à quel point nous sommes capables de relever dans nos listes des marqueurs textuels qui illustrent le bien-fondé de leurs thèses.

Force nous est de constater que la simple lecture des listes-outils est aussi presque toujours surprenante et suggestive: nous y trouvons quelques résultats d'amplitude inattendue. En effet, d'une part on découvre des particularités du texte qui n'étaient jamais apparues à la lecture: tel terme présente une fréquence que nous étions très loin de soupçonner, tel autre se trouve la plupart du temps dans l'entourage immédiat d'un mot bien précis... ; d'autre part on est quelques fois surpris de constater qu'un mot est très peu, voire pas du tout présent dans le texte, ou encore qu'il n'apparaît que dans un contexte précis et prend de ce fait un sens autre que celui que nous escomptions. Bien des raffinements de notre analyse viendront donc des surprises et des déceptions découlant de la lecture de ces listes-outils.

Les résultats purement statistiques relèvent plus de la stylométrie et entrent moins en ligne de compte dans notre stratégie d'analyse. Cependant, j'espère être en mesure d'obtenir le magnétoscrit d'un roman de Novalis, *Heinrich von Ofterdingen*, les résultats de la comparaison statistique n'en prendront alors que plus de relief.

* * *

Il est sûr et certain que le Traitement Automatique du Texte comme phase additionnelle du processus interprétatif apporte beaucoup aux chercheurs en littérature, mais fait aussi ressortir avec peut-être plus de netteté des questions que chacun d'entre eux affronte dans la pratique quotidienne de sa recherche.

— Comment m'assurer de la légitimité et de la validité du lien que j'établis entre le textuel et le conceptuel lors de mes interprétations? À savoir: la constellation de mots/éléments textuels est-elle bien celle qui illustre le thème/concept interprétatif choisi et seulement ce dernier?

— Lors de la sélection du matériau textuel, une part des informations du texte est laissée pour compte. Quel est le statut de ce résidu? — Est-ce qu'une méthode interprétative doit être évaluée à l'aune de la quantité d'éléments textuels qu'elle implique et doit-elle être comparée sous cet angle à d'autres méthodes pour arriver à déterminer laquelle est la plus pertinente pour mon objet d'étude?

Enfin, même si le texte de Novalis présente à forte densité des problèmes de saisie et de repérage d'information, difficulté que beaucoup de textes littéraires ne présentent que de manière plus ténue, je reste cependant de plus en plus convaincu que l'étape informatique du travail interprétatif deviendra non plus un détour, mais le chemin normal, peut-être même le nouveau paradigme de toute analyse littéraire.

Université de Montréal