

Analyse structurale de contenu et soutiens logiciels : une introduction au projet Anaconda

Jean-Pierre Hiernaux, Ph.D.

Université Catholique de Louvain

Résumé

Seront d'abord clarifiés l'objet spécifique, les fondements et les modes opératoires propres à l'analyse structurale de contenu. Au regard des conditions pratiques que celle-ci demande de gérer face à des données concrètes, se raisonne ensuite l'adéquation des types de soutiens logiciels possibles. Enfin seront présentés quelques dispositifs plus particulièrement appropriés, tels que nous les développons et testons actuellement à l'échelle du projet ANACONDA - pour ANALYse de CONtenu Disjonctive et Associative¹ - exploitant les ressources du traitement de texte Microsoft Word sous programmation VBA.

Mots clés

ANALYSE STRUCTURALE, ANALYSE DE CONTENU, LOGICIELS, CAQDAS

L'analyse structurale de contenu - objet, fondements, pratique

L'objet

Devant des manifestations de sens, notamment verbales et discursives, l'objet de l'analyse structurale de contenu consiste à saisir non pas le mode de manifestation du sens, non pas la forme ou la pratique discursive, mais bien le mode d'existence du sens comme tel, c'est-à-dire ce par quoi ce qui est évoqué se constitue en « réalité » perceptible ou concevable tout en se dotant de propriétés spécifiques. Par delà l'échelle discursive, cet objet de connaissance désigne donc l'ordre même du sens, un phénomène proprement sémantique.

Pour le praticien des sciences sociales cependant, pour l'analyste des comportements humains, le phénomène sémantique dont il s'agit ne se réduit pas à l'échelle de la langue comme telle où il renverrait aux « dénotations » des mots et expressions, c'est-à-dire à leurs investissements de sens standard dont rendent compte les définitions des dictionnaires et qui permettent de communiquer en sachant que si je lis ou entends « masculin », cela veut dire

RECHERCHES QUALITATIVES – Hors Série – numéro 9 – pp. 56-82.

LOGICIELS POUR L'ANALYSE QUALITATIVE: INNOVATIONS TECHNIQUES ET SOCIALES

ISSN 1715-8702 - <http://www.recherche-qualitative.qc.ca/Revue.html>

© 2010 Association pour la recherche qualitative

ceci. Alternativement, en effet, c'est sur les « connotations » que l'analyste se centrera en tentant de saisir – par delà ce qui en permet la communication – ce qui compose comme telles les représentations, les « réalités » ou les évidences particulières qu'assument des locuteurs ou des fractions de locuteurs spécifiques, ce qui – en même temps qu'un phénomène sémantique – constitue un fait sociologique, psychologique et/ou historique.

Ainsi, d'une page de « catéchisme » qu'un prêtre catholique, professeur de religion, composa pour ses élèves – il y a plus de 25 ans déjà, mais cependant bien après mai '68 – , une analyse structurale de contenu retiendra non le texte comme tel, ni son style, ni sa rhétorique, ni non plus ce qui fait que quiconque comprenant la langue utilisée comprendra aussi ce que désignent les mots « masculin », « matière », « exploiter » ou tout autre élément évoqué, mais bien le réseau constitutif de sens spécifique et particulier dont rend compte le graphe de la Figure 1.

Dans cette figure, on voit notamment comment – dans les lignes d'abord – les destinées se construisent au nombre de deux seulement et ce en se contre-définissant comme mâles ou femelles, comment aussi – au sein des objets de chacune de ces destinées – la matière se construit et tant qu'inverse de l'humain et réciproquement tandis que les actions des mêmes destinées se spécifient exclusivement et typiquement soit comme exploitation, soit comme valorisation, chacun de ces termes prenant sens comme inverse de l'autre. Par ailleurs, – en colonnes cette fois – l'exploitation se donne comme une action propre aux espaces extérieurs (extra-domestiques, publics), compatible seulement avec les objets d'ordre matériel tandis que la chaîne EXPLOITER – EXTÉRIEUR – MATIÈRE caractériserait la destinée mâle en la disjoignant de l'assortiment inverse où l'humain se valorise dans l'espace intérieur et constituerait ainsi la destinée femelle... Voilà exactement ce qui nous intéresse.

Les fondements

Au fondement de l'analyse structurale de contenu, la sémantique structurale prétend à des concepts et outils descriptifs congruents avec la nature intrinsèque que l'élaboration scientifique reconnaît aux phénomènes à décrire. À cet égard, si intellectuellement les phénomènes quantitatifs s'élaborent parfaitement dans l'ordre de la fréquence et tirent de ce qui régit cet ordre les instruments appropriés à leur traitement, les phénomènes sémantiques de leur côté, donc les phénomènes de sens, tiennent à l'ordre de la différence et de la ressemblance dont, à leur manière, ils peuvent recevoir des modalités de description adéquate. En particulier, l'existence même du sens peut se saisir comme résultant de la mise en oeuvre de différences : sans différencier, en effet, rien n'émerge à la conscience. C'est ce que, sans pour autant être

destinées	
MASCULIN	FÉMININ
objets	
MATIÈRE	HUMAIN
espace	
EXTÉRIEUR	INTÉRIEUR
actions	
EXPLOITER	VALORISER

Figure 1. Exemple d'un réseau constitutif de sens

linguiste mais de manière néanmoins fort séduisante, illustre le romancier Blaise Cendrars (né Frédéric-Louis Sauser, 1887-1961) en annexe de son roman *Moravagine* où il apparaît que la langue martienne ne connaît qu'un seul mot - *areukeukokex* - lequel veut tout dire... et qui donc ne veut rien dire, faute de faire la différence avec quoi que ce soit. A cet égard justement, la sémantique structurale pose comme forme minimale constitutive de sens, la relation de disjonction qui au sein d'une totalité initiale (Ti) fait émerger comme réalités spécifiques un couple d'unités de sens (Ai, Bi) qui se constituent comme telles en se contre-définissant : Ai / Bi (Greimas, 1966). Ainsi, par exemple, les « réalités » que désignent les mots « gauche » et « droite » naissent de la contre-définition qui se donne non seulement à penser mais même et d'abord à expérimenter lorsque l'on étend les bras à l'horizontale au sein de la totalité initiale qu'est à ce moment-là la « latéralité ». Ainsi aussi naissent par disjonction au sein de leurs totalités initiales les couples contre-définis que nous avons observés dans le graphe présenté plus haut. Et il en va de même des milliers d'autres couples du même type qui peuplent notre esprit d'autant de « réalités » et de notions qui orientent notre existence.

Certes, si la forme minimale constitutive de sens est du type que nous venons de situer, toute structure de sens plus large peut se saisir comme l'articulation d'une pluralité de telles formes selon un mode déterminé (Hiernaux, 1978). Ainsi, une articulation du genre évoqué et le mode qui la caractérise apparaissent-ils dans le graphe que nous présentions plus haut : celui-ci en effet doit sa structure typique au fait que lorsqu'un terme d'une disjonction donnée (mettons Ai de Ai / Bi) s'y associe à un terme d'une autre disjonction (mettons à Aj de Aj / Bj), ce qui engendre donc l'association Ai-Aj), les deux termes restants s'associent entre eux (soit Bi-Bj). C'est ce qu'un jargon commode, basé sur l'apparence générale du graphe obtenu, a fini par nommer une « structure en parallèle » (voir Figure 2).

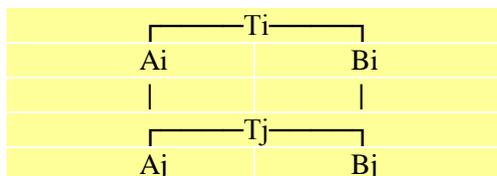


Figure 2. Principe d'articulation des structures en parallèle

Ces structures en parallèle, dont le mode d'articulation est le plus simple, sont aussi les plus fréquentes au sein des manifestations de sens. En même temps, elles ne sont pas les moins décisives quant à l'orientation des comportements.

D'autres modes basiques d'articulation, moins fréquents et que nous ne ferons ici qu'évoquer, peuvent évidemment prendre corps :

- ainsi, lorsqu'au moins deux disjonctions élémentaires interviennent dans un ordre contraint (rang "i" avant "j") et qu'un terme de la première (mettons Bi) forme à son tour la totalité de la seconde (Bi=Tj), apparaît la « structure en éventail » comme ci-dessous :

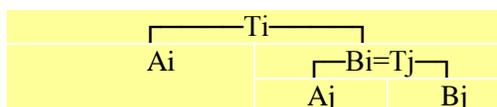


Figure 3. Principe d'articulation des structures en éventail

- ainsi encore, lorsqu'au moins deux disjonctions élémentaires interviennent simultanément comme dans la structure en parallèle mais, cette fois, en réalisant toutes les combinaisons possibles entre leurs termes, apparaît la « structure croisée »² :

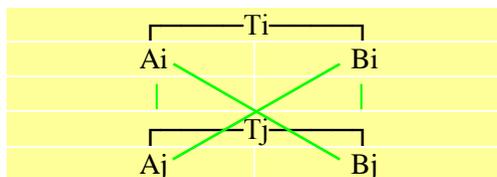


Figure 4. Principe d'articulation des structures croisées

Les graphes et modes d'articulation ci-dessus, faut-il le dire, s'en tiennent aux types de structure basiques c'est-à-dire formellement purs, la

possibilité existant par ailleurs que des manifestations de sens déterminées articulent entre-elles des séquences relevant tantôt de l'un ou de l'autre de ces types. Quoiqu'il en soit, là où il s'agit de rendre compte du mode d'existence du sens, il apparaît combien aux yeux de l'approche proposée le sens est une question de différences et de ressemblances, autrement dit de relations et de rapports qui séparent ou rapprochent, donc de positionnements respectifs qui font les unités en les distribuent les unes par rapport aux autres au sein de ce qui apparaît en substance comme un phénomène topologique³. Ainsi, notamment, l'investissement sémantique des unités de sens tel qu'il se donne à observer, leur sens tel qu'il peut être décrit le plus objectivement possible, se saisit finalement comme résultant de – et équivalant exactement à – la place que ces unités occupent au sein des structures dont elles émergent et telles que permettent de les saisir les limites d'une manifestation déterminée. Par exemple, en posant la question de savoir ce qu'est – dans le cadre de la Figure 1 considérée plus haut – la « matière », la réponse – dans les limites mêmes de sa validité empirique – s'affirme simple et précise : la « matière » est ce qui s'associe à l'« extérieur », à l'« exploitation », à la « destinée masculine » tout en se disjoignant de l'« humain », de l'« intérieur », de la « valorisation » et de la « destinée féminine ». Autrement dit, la définition qu'assumerait un dictionnaire issu de l'analyse des données retrace, au sein de l'ensemble topologique qui la génère, la position du terme qu'elle définit.

La pratique

La pratique de l'analyse structurale de contenu consiste ainsi et en substance à dégager des données concrètes les disjonctions élémentaires, les unités de sens qu'elles constituent et les structures plus larges dans lesquelles les unes et les autres prennent place et corps. Les aboutissements classiques d'une telle pratique sont – au terme d'autant d'esquisses provisoires et de (re)mises au point successives qu'il est nécessaire – les graphes des structures sémantiques qui sont à l'œuvre, soit la représentation topologique des relations qui déploient et articulent les unités de sens d'un même espace logique, d'une même « isotopie », même si les données à leur propos sont dispersées en différents lieux ou moments physiques de manifestations de sens plus ou moins amples.

Ainsi, pour compléter l'illustration, considérons la structure parallèle présentée ci-dessous en forme de tableau et qui résulte d'éléments dispersés sur plusieurs pages du récit de vie d'un immigré maghrébin (Figure 5).

<p>'ULT'A ACCOMPLISSEMENT+ VIE MEILLEURE+ 'ESP'A VILLE+ 'TPS'A PRÉSENT EUROPÉEN+ 'APP'A CONNAISSANCE+ MULTIDIMENSION+ OUVERTURE+ AUTONOMIE+</p>	<p>'ULT'B INACCOMPLISSEMENT- VIE FRUSTE- 'ESP'B CAMPAGNE- 'TPS'B PASSÉ COLONIAL- 'APP'B IGNORANCE- UNILATÉRALITÉ- CLÔTURE- ALIÉNATION-</p>	<p>'ULT'T devenir de soi qualité existentielle 'ESP'T socio-morphologie 'TPS'T socio-chronologie 'APP'T °cognitivité °supplétivité °perméabilité °liberté</p>
--	---	--

Figure 5. Exemple de structure en parallèle sous forme de tableau

On remarquera, du point de vue technique, d'une part que les totalités ('T) des disjonctions ont reçu leur description dans la colonne de droite du tableau, sur la ligne même des couples des contre-définitions correspondantes et, d'autre part, que les disjonctions ont par ailleurs été classées selon les plans de perception qu'elles contribuent à structurer et auxquels correspondent empiriquement une ou plusieurs disjonctions : deux pour le plan des ultimités ('ULT), une respectivement pour les plans de l'espace et du temps ('ESP; 'TPS), quatre pour le plan des appuis ('APP).

D'autre part, du point de vue du contenu, il apparaît notamment comment le locuteur met en jeu une quête dont les ultimités ('ULT'A / 'ULT'B) se raisonnent en termes de devenir de soi et de qualité existentielle tout en liant à des alternatives spatio-temporelles où le milieu urbain ('ESP'A) et le présent européenisé ('TPS'A) s'arrachent comme polarités positives à la campagne ('ESP'B) et au passé colonial ('TPS'B), chacun des binômes ainsi formés s'articulant du reste avec les appuis positifs ('APP'A) ou négatifs ('APP'B) à la quête poursuivie, lesquels se contrastent en termes de développement cognitif, de richesse des apports, de perméabilité sociale et culturelle et enfin d'autonomie des individus... soit toute une perception du monde articulée à l'histoire et au devenir de soi au sein de ce monde.

Ainsi encore les structures parallèles comparées ci-dessous, issues cette fois de centaines de pages de données reprenant d'une part les retranscriptions de débats de groupe entre locuteurs d'âge mûr dans des villages traditionnels d'il y a bientôt un demi-siècle et, d'autre part, des manifestations de sens diverses (livres, articles, tracts, publicités, manifestes, déclarations...) collectées au sein de milieux urbains « dans le vent » avant la fin de l'expansion économique des années '70 (Figure 6).

SOI		SOI	
RÉGULATION+	IRRÉGULATION-	COMPLÉTUDE+	INCOMPLÉTUDE-
ACTIONS		ACTIONS	
S'EFFORCER+	SE LAISSER ALLER-	S'AFFIRMER+	SE SOUMETTRE-
MOYENS		MOYENS	
CONTRAINTE+	LIBERTÉ-FACILITÉ-	LIBERTÉ-FACILITÉ+	CONTRAINTE-
SOURCES		SOURCES	
CONSTRICTEURS+	DÉCONSTRICTEURS-	DÉCONSTRICTEURS+	CONSTRICTEURS-
ULTIMITÉS		ULTIMITÉS	
ORDRE SOCIAL+	CHAOS-	BIEN-ÊTRE+ BONHEUR+	MISÈRE- SOUFFRANCE-
ESPACE		ESPACE	
INTÉRIEUR SOCIAL+	EXTÉRIEUR SOCIAL-	EXTÉRIEUR SOCIAL+	INTÉRIEUR SOCIAL-
TEMPS		TEMPS	
PASSÉ- CONTINUÉ+	FUTUR DE CHANGEMENT-	PRÉSENT et/ou FUTUR+	PASSÉ-

Figure 6. Exemple de comparaison de structures en parallèle

Du point de vue technique, l'on remarquera l'usage d'une grille de plans commune permettant des comparaisons sur base standard. Du point de vue du contenu, apparaît alors le contrepoint entre d'une part, une vision de soi et du monde de type traditionnel-ascétique et d'autre part, une vision de promotion individuelle et sociale de type hédoniste. Selon l'une, doit se conquérir la régulation de soi par l'effort que permet la contrainte qu'offrent divers « constricteurs » (les autorités et tout ce qui nous résiste) de manière à garantir l'ordre social tandis que l'intérieur socioculturel et la tradition sont les espaces-temps élus. Selon l'autre, motivera la complétion de soi à tous égards en s'affirmant grâce à la liberté et à la facilité qu'offrent autant de « dé-constricteurs » (politiques, psycho-socio-pédagogiques, industriels et commerciaux...), l'enjeu étant une ultimité de bien-être et de bonheur tandis que l'extérieur socioculturel et le changement actuel et futur forment, au regard de désirs insatiables, les réservoirs inépuisables de tous les possibles. On remarquera également comment, en passant d'un scénario à l'autre, des éléments objectivement identiques changent de position, donc de sens et de rôle : ainsi s'inversent les statuts de la contrainte et de la liberté, des constricteurs et des dé-constricteurs, de l'intérieur et de l'extérieur, du passé et du futur; ainsi aussi, changent non seulement de statut mais même de nom des éléments tels « s'efforcer » que la négativisation traduit en « se soumettre » (s'efforcer, c'est évidemment se soumettre à une exigence...) ou « se laisser aller » que la positivisation cette fois traduit en « s'affirmer » (qui, en effet,

consiste pour le moins à ne pas renoncer à notre cours spontané...). Voilà aussi que la parure des mots s'assortit aux perspectives qui s'en drapent!

Face aux données et aux soutiens logiciels

De masses de données emmêlées l'analyste fait ainsi surgir des graphes ou tableaux ordonnés, condensés et structurés. Mais entre la tête et la queue du processus, quelle est la chair et quelle est l'alchimie? quelle est la systémativité de la démarche? quelle est la complétude de l'analyse? quelles sont la vérifiabilité et la transférabilité des descriptions? peut-on en retracer la genèse autrement que par l'énoncé formel de ce qu'elle aurait dû être? et la pratique analytique pourrait-elle se partager avec ceux qui, comme Saint Thomas, aimeraient mettre les mains dans les plaies du Seigneur? ou encore pourrait-elle être transférée à d'autres qui la poursuivraient à partir du stade déjà atteint? Ou bien serions-nous plutôt dans le règne de la divine inspiration, l'artiste, sans rien laisser transparaître d'un quelconque procédé, tirant lestement du chapeau le blanc lapin qui ébahit les foules? voire serions-nous devant de plus obscurs tours de passe-passe dont l'un ou l'autre s'illustrerait marginalement d'extraits succulents sélectionnés pour le besoin de la cause et pour l'effet escompté?

Des alternatives positives, sans doute, pourraient tirer profit de quelque utile soutien logiciel. Mais, s'agissant de décrire des relations et des positions, les outils à base fréquentielle – et par là volontiers automatisables – relèveraient en même temps du non-sens : ni le sens ni la topologie, justement, ne tiennent à la fréquence. Quant aux phénomènes pertinents, l'automatisation de leur description demeure largement hors portée tant elle rencontre des obstacles difficilement surmontables. Ainsi en va-t-il de la complexité voire de la subtilité des modes de manifestation des relations entre unités de sens : si un discours peut énoncer directement et clairement que quelque-chose est à l'opposé d'une autre ou, au contraire, s'associe à celle-ci, une infinité d'autres modalités manifestent des relations assumées comme évidentes... même et y compris lorsque les locuteurs – pas plus qu'ils ne sont conscients de la grammaire qu'ils pratiquent – ignorent les rapports qu'ils assument tout en les manifestant néanmoins. Ainsi en va-t-il aussi des manifestations polymorphes des unités de sens, lesquelles peuvent, selon le cas, correspondre à un mot ou à autant de synonymes, à une locution ou à plusieurs et parfois à beaucoup, voire à des séquences plus amples ou à d'autres modalités encore. Ainsi encore le fait que la description des unités de sens ne va jamais sans requérir quelque abstraction pour transcender les variations non-pertinentes au regard du propos de l'analyse tandis qu'en même temps des recherches différentes impliquent des niveaux de pertinence et donc d'abstraction variables servis par la production de vocabulaires descriptifs assortis. Le dernier tableau présenté plus

haut en livre un exemple avec les termes « CONSTRICTEURS » et « DÉCONSTRICTEURS », où le premier vaut pour toute source de contrainte ou de limitations et le second pour toute source de licence ou de facilités. Sans cette condensation-réduction descriptive, qui abstrait à un degré relativement élevé, il n'eut jamais été possible de rendre compte du rapport d'inversion que les deux perspectives décrites présentent sur le plan des « SOURCES », rapport qui n'apparaît évidemment que lorsque se saisit sa base commune. Or, les termes concrets qui apparaissent à l'échelle des manifestations respectives sont, d'un côté, la religion comme police des mœurs, Dieu comme source de la règle, le prêtre, les enseignants et les parents comme surveillants de la conformité, les contraintes mêmes de l'existence, celles de la nature, etc. et, de l'autre côté, le progrès, le socialisme, le Concile du Vatican, Che Guevara, les psychothérapeutes et les gourous, les pédagogues libertaires, les producteurs de biens et de services, etc. Ce n'est donc que lorsque se transcendent toutes les différences non pertinentes du point de vue considéré qu'apparaissent dans les collections ci-dessus les convergences pertinentes que tente d'exprimer le vocabulaire descriptif créé ad hoc à l'aboutissement d'un processus d'observation et en même temps d'élaboration intellectuelle auquel – à l'heure actuelle à tout le moins – l'on voit mal un automate s'atteler et, surtout, aboutir avec quelque pertinence.

Les soutiens logiciels appropriés se réduisent donc modestement à ceux où :

- les instruments essentiels de l'analyste demeurent ses propres yeux et cerveau;
- cet analyste ébauche, en autant d'esquisses et de (re)mises au point successives – tels des escarpins essayés aux pieds des données – les graphes des structures qu'il voit émerger, ce qui confirme qu'en matière d'analyse structurale de contenu comme en bien d'autres domaines la rigueur des instruments, conceptuels ou matériels, n'ôte rien à l'art que demeure leur mise en œuvre optimale ni à ce que doit à cet art le résultat obtenu;
- s'offre le soutien d'un certain nombre de commodités marginales qui, tirant profit des disponibilités informatiques contemporaines, remplacent avantageusement les crayons, les surligneurs, les ciseaux, la colle et les boîtes à chaussures qui depuis belle lurette déjà aidaient les analystes à identifier, collecter et classer les traces de leurs objets de connaissance et les pièces à conviction de leurs démonstrations.

Au travers des énoncés qui précèdent, on devrait avoir reconnu les types de soutiens logiciels affublés de l'inélégant acronyme CAQDAS, qui – en une

langue étrangère mais dominante – vaut pour *Computer-Assisted Qualitative Data Analysis Software*, longue locution dénotant comme telle peu de choses spécifiques⁴ mais qui vaut là où, alternativement à la capacité qu’ont les ordinateurs de compter, s’exploitent plutôt celle qu’ils ont aussi à agir sur du texte : annoter, sélectionner, copier, coller, collecter, comparer, classer, etc. Des types de soutien donc auxquels devrait intéresser pour notre cas moins la raison – ou la mode? – des nœuds et arborescences⁵ que la contribution à la restitution des relations et topologies sous-jacentes aux données.

Le projet Anaconda – dispositifs de base et détails, compléments et technique

Les dispositifs de base

Les soutiens logiciels aux analyses structurales de contenu tels que testés sous le projet Anaconda adaptent des procédures relativement simples proposées déjà antérieurement pour le traitement manuel de corpus de données volumineux (Hiernaux, 1995). Là où l’élaboration des données consiste essentiellement à dépasser la dysmorphie entre d’une part la manifestation de sens observée et d’autre part l’organisation structurale sous-jacente, ces procédures et modalités contribuent à trois types d’opérations distribuées à leur tour sur trois espaces de travail :

- d’une part, sur l’espace de travail initial que forment les données à traiter (texte, retranscription...), les opérations de classement qui, par delà l’entremêlement des données à l’échelle de la manifestation, vont en rassembler les segments jugés pertinents dans des collections davantage homogènes;
- d’autre part, sur l’espace de travail que forment à leur tour les collections obtenues, les opérations de description qui, par delà la variété des modes de manifestation des contenus collectés, vont en condenser et énoncer au niveau d’abstraction jugé opportun les unités de sens et leurs disjonctions constitutives;
- enfin, sur un dernier espace de travail, l’esquisse de la structure d’ensemble qui, jusqu’à sa mise au point finale, va se dégager progressivement des relations observées entre les différents couples disjonctifs qui auront été décrits.

Le dispositif de base d’Anaconda vise à articuler les trois espaces de travail concernés en un ensemble synoptique et interactif où notamment :

- ce qui dérive des opérations réalisées dans l’un des espaces et qui importe à l’un ou au deux des autres y est automatiquement répercuté et visualisé;

- depuis les éléments présents dans un espace donné, (i) il est possible de retracer la voie vers les espaces en amont dont ces éléments procèdent et notamment d'y reprendre, poursuivre ou amender les opérations correspondantes, (ii) à terme, il est possible également, de re-visionner la progression des élaborations vers l'aval.

Techniquement, ce qui vient d'être évoqué peut être réalisé de différentes façons. Anaconda, pour sa part, a opté pour offrir comme support aux espaces de travail et à leur articulation de simples fichiers de traitement de texte : un par espace de travail. Gérés par le logiciel, mais lisibles, imprimables, échangeables... même hors celui-ci, ces fichiers – sauf évidemment celui qui contient les données en l'état initial – se créent et se dérivent automatiquement les uns des autres, s'affichent de même au moment opportun – en plein écran pour la commodité – et sont tout aussi automatiquement coordonnés quant à leurs contenus, noms, enregistrements, ouvertures, suppressions, protections, etc.⁶

Chacun des trois espaces de travail illustrés dans l'articulation de la Figure 7 donne ainsi exemple de l'affichage sur écran d'une partie de l'un des fichiers correspondants... même s'il a fallu ici, vu la surface réduite, raccourcir singulièrement les éléments illustrant les contenus.

Sur base de l'illustration ci-dessous, le fonctionnement du dispositif peut se détailler davantage dans ce qui suit.

(1) Au sein de l'espace/fichier 1, celui des données à traiter, l'analyste peut :

- par des procédures familières aux traitements de texte, sélectionner tout segment désiré depuis un seul mot jusqu'à un paragraphe entier;
- insérer en tête de chaque sélection au moins une marque d'indexation, une étiquette, qui détermine à quelle collection se joindra, dans l'espace/fichier des collections d'extraits et comme davantage explicité plus bas, un extrait sous forme de copie de la sélection concernée.

Dès qu'une sélection est réalisée et indexée, Anaconda le confirme par un surligné jaune. En même temps, dès la première sélection et indexation, les données initiales sont automatiquement réenregistrées dans un nouveau fichier – le « fichier indexé » – qui se substitue dans l'espace de travail en cours au fichier contenant les données vierges, lequel aura été fermé intact et sera conservé comme tel à toutes fins ultérieures.

Les marques d'indexation, pour leur part, répondent au type de structure

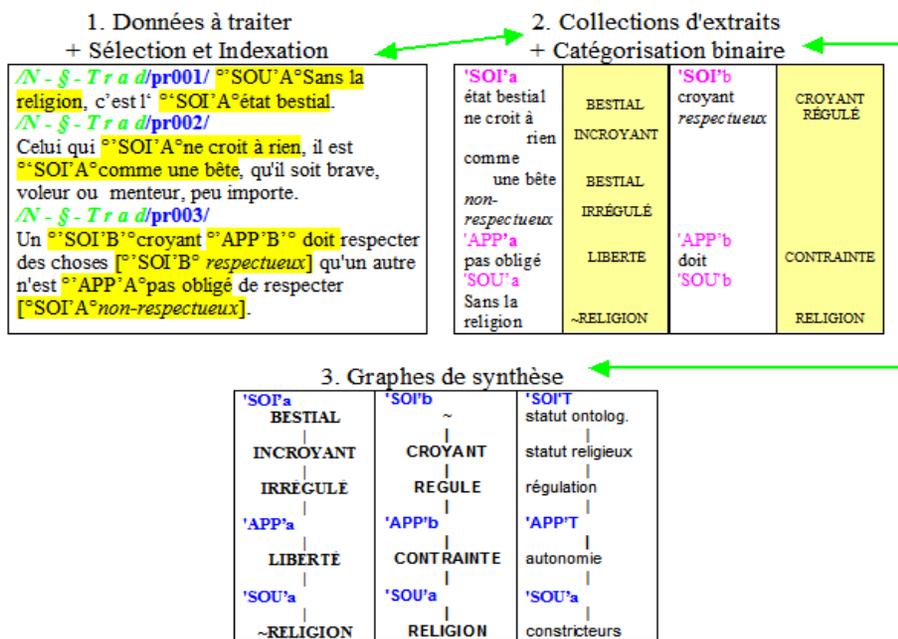


Figure 7. Les trois espaces de travail articulés d'Anaconda

auquel l'analyste pressent que participent les données qu'il sélectionne. Ainsi, en Figure 7 plus haut, espace/fichier 1, on verra l'exemple d'un fichier indexé à l'échelle de sélections qui préfigurent une structure en parallèle, type de structure auquel nous nous tiendrons dans ce qui suit⁷. Les marques d'indexation se composent dès lors des éléments suivants :

- un radical formé du sigle que l'analyste crée pour désigner le plan de perception du réel dont il voit manifester un contenu dans la sélection indexée;
- un suffixe 'A ou 'B notant la polarité du contenu concerné au sein d'une disjonction binaire du plan considéré.

Repris de la Figure 7 plus haut, espace/fichier 1, les exemples du Tableau 1 illustrent et explicitent davantage l'usage de telles marques d'indexation⁸.

Il importe de noter que les radicaux des marques d'indexation du type ici concerné n'équivalent pas aux totalités T des disjonctions binaires élémentaires, comme le serait par exemple la « latéralité » pour la disjonction gauche/droite. Ils réfèrent au contraire à des plans de perception du réel qui peuvent contenir plus d'une disjonction binaire : ainsi, en Tableau 1 ci-dessous,

Tableau 1
 Marques et pratiques d'indexation - exemples

°°SOI'A°état bestial	L'analyste voit dans la séquence /état bestial/ une manifestation du plan des « états de soi »; il crée donc le sigle 'SOI et l'insère en tête de la sélection, suivi de la mention de sa polarité : 'A ⁹ .
°°SOI'A°ne croit à rien	L'analyste voit /ne croit à rien/, i.e. « incroyant », manifester également le plan des « états de soi » avec la même polarité que ci-dessus : 'A.
°°SOU'A°Sans la religion	L'analyste voit /Sans la religion/ manifester le plan des « sources des états de soi »; il crée le sigle 'SOU et - /Sans la religion/ s'associant à /état bestial/ ('A) - il l'insère avec la polarité 'A.
°°SOI'B°croyant	L'analyste voit /croyant/ manifester la polarité inverse à /ne croit à rien/ ('A) plus haut ; il insère donc le sigle 'SOI suivi de la polarité 'B.
°°APP'B°doit	L'analyste voit /doit/, i.e. la contrainte, manifester le plan des « appuis aux états de soi »; il crée le sigle 'APP et l'insère avec la polarité 'B : /doit/ s'associe en effet à /croyant/ ('B).

on pourrait – en anticipant sur des étapes d'élaboration ultérieures – voir le plan 'SOI annoncer déjà deux disjonctions binaires BESTIAL / NON-BESTIAL et INCROYANT / CROYANT. Ce caractère relativement englobant des marques d'indexation est essentiel. En effet, confronté aux données sous la forme où elles se donnent, l'analyste n'est dans bien des cas pas à même ni de saisir ni d'énoncer d'emblée les unités de sens impliquées et encore moins leurs disjonctions constitutives. Souvent, il ne pourra le faire qu'après qu'aient été collectées un nombre suffisant de manifestations qui, sous des formes diverses, éclairent ces unités et disjonctions. Dans l'intervalle, et pour ouvrir à des progressions ultérieures, l'analyste doit donc rassembler les données en collections compatibles avec les possibilités d'observation qu'il peut actuellement pratiquer, soit celles en termes de regroupements encore relativement larges. De tels regroupements ne se ramènent cependant pas à des pis-aller : le nombre et la variété relative des manifestations qu'ils peuvent contenir offriront aussi les bases les plus favorables pour des descriptions ultérieures dont les niveaux d'abstraction alternatifs ne pourront apparaître qu'à ces conditions et à ce moment-là¹⁰. Enfin, lorsque des descriptions plus fines deviendront possibles – telle celle des disjonctions élémentaires –, leurs

résultats, plutôt que d'apparaître en vrac, prendront place au sein de tels regroupements et bénéficieront ainsi du classement de rang supérieur que commencent à esquisser, sous forme de plans, les radicaux des marques d'indexation. – Voilà qui non seulement aura situé au niveau des procédures d'indexation une condition de production des collections d'extraits, mais encore qui aura suggéré l'importance de telles collections à différents égards et pour différents moments du traitement des données.

(2) Dès que l'analyste réalise une sélection et indexation dans l'espace/fichier des données à traiter, l'espace/fichier des collections d'extraits (i) est automatiquement créé s'il n'existe pas encore et (ii) en tous cas – et tant que l'analyste ne décide pas de revenir à l'espace/fichier antérieur ou d'en rejoindre un autre – est affiché en visualisant l'extrait qui vient d'être inséré au sein de la collection à laquelle ses marques d'indexation l'ont destiné, collection qui, si elle n'existait pas encore, aura elle aussi été simultanément et automatiquement créée.

Au détail :

- l'espace/fichier des collections d'extraits anticipe sous forme de tableau le modèle du type de structure qu'il contribue à traiter : ainsi, en Figure 7 plus haut, le modèle de l'espace/fichier 2 est celui des structures en parallèle;
- horizontalement, le tableau pour structures en parallèle s'organise en sections transversales selon les plans de perception dont l'analyste a, lors des opérations de sélection et d'indexation dans l'espace/fichier des données à traiter, constaté la manifestation et créé les sigles qu'il a ensuite insérés aux marques d'indexation de ses sélections : dès qu'au moins une de ces sélections reçoit une marque d'indexation contenant le sigle d'un plan donné, ce sigle est automatiquement reporté dans l'espace/fichier des collections d'extraits et y définit une section horizontale qui pourra recevoir les extraits correspondants; – ainsi, en Figure 7 plus haut, espace/fichier 2, les sections/plans apparaissent en rose et traversent toutes les colonnes du tableau respectivement pour 'SOI (« états du soi »), 'APP (« appuis aux états du soi ») ou 'SOU (« sources des appuis »);
- verticalement, le même tableau s'organise en colonnes qui divisent les sections ci-dessus en parties gauche ou droite correspondant aux polarités 'A ou 'B des marques d'indexation que l'analyste attribue aux sélections réalisées dans l'espace/fichier des données à traiter et dont dérivent les extraits; – en fait, comme l'illustre la Figure 7 plus haut, espace/fichier 2, les espaces-colonne gauche et droite sont encore une

fois divisés en deux : les extraits s'insèrent plus précisément en colonne 1 ou 3 tandis que les colonnes 2 et 4 restent libres pour des traitements que nous préciserons bientôt;

- sur le canevas résultant des paramètres ci-dessus, et en fonction des marques d'indexation attribuées à la sélection dont il procède, chaque extrait rejoint donc automatiquement l'espace d'une collection qui correspond, en ligne, à un plan de perception donné et, en colonne, à une polarité donnée; – à ce stade, notons-le, il n'est pas visé à ce que, au sein de la collection des extraits d'un plan de perception donné, plan auquel peuvent appartenir plusieurs totalités et leurs disjonctions respectives¹¹, les extraits 'A relevant de la disjonction d'une totalité donnée se trouvent nécessairement en face d'extraits 'B relevant de la même totalité : comme déjà évoqué, il est en effet et fréquemment impossible de garantir de telles identifications lors des sélections et indexations dont dérivent tant les extraits que leur classement tandis que, comme il sera montré à l'instant, dès que la possibilité de ces identifications survient, elle ouvre aux traitements ultérieurs que l'étape actuelle a justement comme but de préparer.

Automatiquement et progressivement composé, développé et tenu à jour sous l'effet des sélections et indexations que l'analyste réalise dans l'espace/fichier des données à traiter, l'espace/fichier des collections d'extraits est cependant et également le lieu d'opérations subséquentes dès que l'analyste, au terme de – ou même pendant – la production des collections, voit émerger la possibilité de condenser en un énoncé descriptif adéquat les unités de sens, voire même les couples disjonctifs, que des extraits déterminés manifestent¹². L'analyste peut alors en effet, dans l'espace/fichier des collections d'extraits :

- en face d'un ou de plusieurs extraits, insérer une catégorisation, un vocable qui – à cette étape de progression de l'analyse – semble convenir pour condenser l'unité de sens sous-jacente aux contenus concernés : ainsi en Figure 7 plus haut, espace/fichier 2, les catégorisations créées et insérées en colonne 2 pour les extraits de la colonne 1, dont notamment BESTIAL pour les extraits « état bestial » et « comme une bête », ou les catégorisations créées et insérées en colonne 4 pour les extraits de la colonne 3¹³;
- en face d'un ou de plusieurs extraits relevant de polarités opposées au sein d'une même collection, condenser l'énoncé complet de la disjonction sous-jacente sous forme d'une catégorisation binaire enregistrée comme telle par le logiciel, et en insérer les termes respectifs au regard des extraits correspondants : ainsi, en Figure 7 plus haut, espace/fichier 2, au plan 'APP, l'enregistrement comme tel du couple

LIBERTÉ / CONTRAINTE pour les extraits « pas obligé » et « doit » ou, au plan 'SOI, l'enregistrement comme tel du couple INCROYANT / CROYANT même si les extraits correspondants n'étaient pas – et donc si ces termes eux-mêmes ne sont pas – face à face dans la collection, situation que géreront notamment les fonctionnements présentés dans ce qui suit.

(3) Dès que l'analyste insère une catégorisation en colonne 2 ou 4 de l'espace/fichier des collections d'extraits, l'espace/fichier des graphes de synthèse (i) est automatiquement créé s'il n'existe pas encore et (ii) en tous cas – et tant que l'analyste ne décide pas de revenir à l'un ou l'autre des espace/fichiers antérieurs – est affiché en visualisant le plus récent état de l'esquisse de structure d'ensemble qui résulte des traitements jusque là appliqués aux données.

Au détail :

- l'espace/fichier du graphe de synthèse reprend automatiquement le modèle du type de structure que contribue à traiter l'espace/fichier des collections d'extraits et que préfiguraient déjà les marques d'indexation insérées aux sélections dans l'espace/fichier des données à traiter;
- s'agissant de structures parallèles, les deux premières colonnes de l'espace/fichier du graphe de synthèse reprennent ainsi la grille d'intitulés par plans et polarités que tient constamment à jour l'espace/fichier des collections d'extraits qui répercute lui-même l'état des sélections et indexations réalisées dans l'espace/fichier des données à traiter; – une troisième colonne s'y ajoute où l'analyste pourra, comme précisé plus bas, insérer des suppléments d'élaboration; – en Figure 7 plus haut, espace/fichier 3, on verra les intitulés concernés figurer en bleu et l'on y reconnaîtra notamment ceux qui figuraient déjà en rose dans l'espace/fichier 2;
- sur le canevas de plans et polarités qu'offrent ainsi les deux premières colonnes de l'espace/fichier du graphe de synthèse, et selon les coordonnées des extraits concernés dans l'espace/fichier des collections d'extraits, Anaconda va cette fois automatiquement et à chaque attribution d'une catégorisation à un extrait (i) reporter uniquement les catégorisations telles qu'elles condensent et expriment les contenus des extraits sous forme d'unités de sens, et ce sans doublets¹⁴ et (ii) lorsque de telles catégorisations ont été mémorisées comme constituant des couples disjonctifs, en aligner les termes en face à face dans le graphe¹⁵; – l'analyste pourra, s'il le souhaite et comme illustré en Figure 7,

espace/fichier 3, compléter dans la troisième colonne sa description des totalités qui correspondent aux couples disjonctifs;

- comme les positions que vont occuper les catégorisations dans l'espace/fichier du graphe de synthèse correspondent à celles de l'espace/fichier des collections d'extraits et comme à leur tour celles-ci préfiguraient – déjà depuis les sélections et indexations dans l'espace/fichier initial – une forme de structure déterminée, c'est bien le contenu de la structure correspondant à cette forme en même temps que sous-jacente aux données retenues qui émerge au fur et à mesure dans le graphe de synthèse, avec un degré de dépouillement, de netteté et de confirmation empirique accru à chaque fois que, par delà la multiplicité des extraits collectés et des formes de manifestation qu'ils recèlent, davantage d'éléments se traduisent dans les vocables condensés des catégorisations.

(4) Au travers de l'ensemble du dispositif, et passant par deux échelles articulées qui confrontent à chaque fois l'analyste au résultat de chacune de ses opérations – d'une part les sélections et indexations répercutées dans les collections d'extraits et, d'autre part, les catégorisations répercutées dans le graphe de synthèse – s'accomplit ainsi l'objectif même de la description structurale : transcender la masse entremêlée et polymorphe des données initiales pour dégager les structures de sens sous-jacentes et les restituer en graphes synthétiques, condensés et ordonnés en même temps qu'empiriquement justifiés c'est-à-dire – le logiciel aidant ici encore à tenir l'exigence – des graphes finaux dont n'importe quelle catégorisation donne, par la commande adéquate, accès aux collections d'extraits qui la justifient tandis que, de la même façon, chacun des extraits de ces collections livre à son tour l'accès au contexte dont il est issu au sein des données traitées, données qui de la sorte, pour avoir conduit au graphe final, se rejoignent également à partir de lui et sans solution de continuité.

Notons que l'articulation des deux échelles rappelées plus haut ne constitue en rien une séquence aux temporalités contraintes. Tout au contraire, l'analyste peut à son gré et/ou selon les opportunités qu'il discerne lors du traitement d'un lot de données :

- tantôt privilégier, dans une première étape, la réalisation des sélections et indexations, en réservant pour une étape ultérieure la tâche de condensation et de catégorisation descriptive;
- tantôt, chaque fois que l'opportunité en apparaît – et même avant que ne s'atteigne le terme d'une collection –, privilégier la progression jusqu'à

l'attribution de catégorisations aux extraits collectés et donc jusqu'à la contribution à l'esquisse du graphe de synthèse.

Soit, à l'échelle des trois espaces/fichiers articulés, les usages soit séquentiel soit directement transversal schématisés comme le montre la Figure 8.

En même temps, la navigabilité en marche-avant et en marche-arrière au sein du dispositif, jointe à l'articulation de ses différents niveaux, contribue à sa plasticité permanente aux corrections et ajustements : ainsi, partant de n'importe quel élément en aval du processus d'élaboration, n'importe quel élément en amont peut être rejoint et amendé, supprimé ou remplacé, avec répercussion automatique des implications dans n'importe quel espace qui en dérive¹⁶. Concevant du reste que l'analyse – l'essayage des escarpins aux pieds des données – est dans son principe une pratique d'affinement par esquisses provisoires et mises au point successives, Anaconda offre également – d'une part dans l'espace/fichier des données à traiter, d'autre part dans celui des collections d'extraits – des dispositifs de recherche/remplacement susceptibles de gérer notamment des adaptations/modifications en série portant respectivement sur des indexations ou des catégorisations, dispositifs qui eux aussi gèrent, aux différentes échelles concernées, toutes les répercussions impliquées aux niveaux qui en dérivent.

Enfin, lors de l'enregistrement des modifications progressives aux trois fichiers qui servent de support à chacun des espaces de travail décrits jusqu'à présent, Anaconda offre systématiquement l'option de conserver et d'archiver ou non les enregistrements des états antérieurs. Il est ainsi possible, au titre de contribution directe à la pratique analytique ou encore à des fins didactiques – par exemple la révision du mode de travail suivi par un analyste en formation –, de retracer les étapes du traitement des données tout autant que les éventuelles modifications, reprises et réorientations appliquées à ce traitement lui-même.

Quelques détails et compléments

À l'échelle des détails opératoires, les dispositifs de base tels que décrits jusqu'à présent, bénéficient d'un certain nombre de modalités ou de fonctionnalités spécifiquement assorties.

Ainsi, pour les processus de sélection et d'indexation au sein des données à traiter, Anaconda offre notamment les latitudes et appuis suivants :

- la possibilité de multi-indexation, permettant qu'une sélection insécable où l'analyste voit se manifester des contenus relevant de plus d'un plan de perception puisse donner lieu à autant d'extraits-doubles que de plans concernés, chaque collection correspondante se complétant automatique-

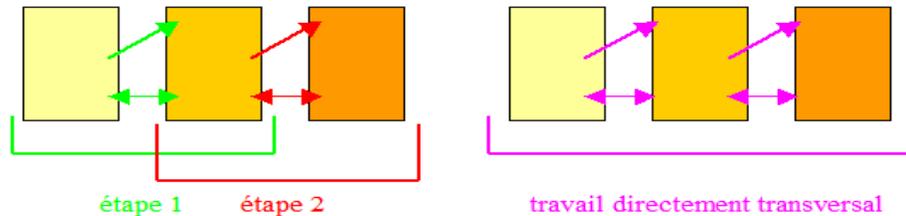


Figure 8. Usage séquentiel ou directement transversal des espaces/fichiers

ment par l'un de ces extraits; – soit, référant aux données traitées en Figure 7 plus haut, espace/fichier 1, paragraphe « pr003 », la modalité alternative d'indexation présentée au Tableau 2 ;

- la possibilité d'éditer les données initiales, permettant par exemple à l'analyste d'insérer au regard de l'original des « traductions » ou des clarifications qui pourront également entrer dans les sélections indexées, ou encore permettant – alternativement à la pratique illustrée en Tableau 2 ci-dessous – de dédoubler des sélections insécables par des copiés-collés de celles-ci ou de parties de celles-ci dans le fichier-même des données; – on trouvera ainsi, en Figure 7 plus haut, espace/fichier 1, paragraphe « pr003 », entre crochets, des exemples de « traduction »;
- la tenue à jour¹⁷ et la mémorisation automatiques des marques d'indexation dans des listes stockées sous forme de fichiers en même temps qu'affichées dans l'espace de travail correspondant d'Anaconda, favorisant ainsi, si l'analyste le souhaite, l'usage de listes-standard pour le traitement de divers lots de données et permettant en tous cas – lors des sélections – l'insertion des marques d'indexation par un simple clic.

À l'échelle des collections d'extraits et des opérations de catégorisation de leurs contenus, par ailleurs, Anaconda offre notamment les modalités et appuis suivants :

- l'adjonction à tout extrait de la référence à la localisation de la sélection dont il procède au sein des données traitées et, s'il y a lieu, de l'information sur la présence simultanée de l'extrait dans plus d'une collection (cas des multi-indexations)¹⁸;
- une tutelle automatique des pratiques de l'analyste lors de la formulation et de l'usage des termes des couples de catégorisations disjonctives, de manière à garantir et/ou gérer la cohérence du vocabulaire descriptif¹⁹;

Tableau 2
Multi-indexation - exemple

<p>°'SOI'A°°'APP'A°pas obligé</p>	<p>L'analyste voit /pas obligé/ manifester un « état de soi » : celui où on est sans contraintes; il insère donc la marque d'indexation 'SOI.</p> <p>L'analyste voit aussi /pas obligé/, i.e. l'absence de contrainte, manifester un appui à un « état de soi » : il insère donc également la marque d'indexation 'APP.</p> <p>L'extrait correspondant ira à la collection 'SOI et aussi à la collection 'APP.</p>
-----------------------------------	--

- de manière analogue à ce qui se passe pour les marques d'indexation, la tenue à jour et la mémorisation automatiques des marques de catégorisation et de leurs couples disjonctifs dans des listes stockées sous forme de fichiers en même temps qu'affichées dans l'espace de travail correspondant d'Anaconda, favorisant ainsi, si l'analyste le souhaite, l'usage de listes-standard pour le traitement de diverses collections d'extraits et permettant en tous cas – lors des opérations de catégorisation – l'insertion des marques correspondantes par un simple clic;
- selon les besoins de l'analyste, le reclassement automatique des extraits au sein des collections d'extraits de manière à favoriser leur consultation soit dans l'ordre historique de la réalisation des sélections et indexations, soit dans l'ordre physique de leur apparition dans les données traitées, soit en fonction de leur traitement en termes de catégorisations; – la dernière modalité citée correspond au fait que, entrés dans une collection dans l'ordre même de leur obtention, les extraits s'y entremêlent volontiers du point de vue des unités de sens que l'analyste pourra y reconnaître lors des opérations de catégorisation; or, pour soutenir ces opérations, il est avantageux que d'une part l'analyste puisse consulter synoptiquement les extraits qui ont déjà reçu une catégorisation déterminée tandis que, d'autre part, ceux qui restent à catégoriser seront eux-mêmes regroupés; c'est l'effet produit comme illustré à la Figure 9 où – passant du tableau de gauche à celui de droite –, les collections d'extraits respectivement 'A (colonnes 1 et 2) ou 'B (colonnes 3 et 4), au départ non autrement classées, sont réorganisées en fonction de l'état de

SOF'a état bestial	BÊTE	SOF'b respectueux	RÉGULÉ
non-respectueux	IRRÉGULÉ	croyant	
ne croit à rien			
comme une bête	BÊTE		

SOF'a ne croit à rien		SOF'b croyant	
état bestial	BÊTE		
comme une bête	BÊTE		
non-respectueux	IRRÉGULÉ	respectueux	RÉGULÉ

Figure 9. Reclassement automatique des extraits - exemple

catégorisation de leurs extraits pour présenter d'abord ceux de ces extraits qui restent à catégoriser (cf. cellules vacantes respectivement en colonne 2 ou 4), puis ceux qui le sont déjà et qui apparaissent à leur tour regroupés selon les marques de catégorisation qui leur correspondent, marques elles-mêmes classées en ordre alphabétique²⁰.

Outre les modalités et fonctionnalités ponctuelles du type qui précède, Anaconda peut ou pourra mobiliser également des dispositifs complémentaires à son fonctionnement de base. Ainsi :

- est disponible un dispositif de normalisation du texte des données et de référencement automatique de ses paragraphes, susceptible de traiter soit un seul fichier de données initiales à la fois, soit une série complète de tels fichiers – par exemple une collection de retranscriptions d'entretiens –, et qui réalise dès le départ les deux opérations automatiques que sont (i) d'une part la standardisation de la présentation des données et la vérification/adaptation de certaines de leurs formes aux modalités des traitements ultérieurs²¹ et, (ii) d'autre part, la numérotation automatique des paragraphes des données, insérant également à chacun de ceux-ci la référence au fichier concerné²²;
- est disponible également un dispositif extracteur automatique de segments de texte, susceptible lui aussi de traiter soit un seul fichier de données initiales à la fois soit une série complète de tels fichiers, et qui réunit, à partir de ces sources et sous forme de copies, des collections de segments-standard – par exemple des paragraphes – où, au gré de l'analyste, Anaconda reconnaît (i) soit la présence d'une chaîne de

caractères déterminée (un mot, une chaîne de mots, un lemme...), (ii) soit la présence d'au moins une chaîne de caractères déterminée parmi un ensemble de telles chaînes défini par l'analyste (par exemple en tant qu'indicateurs d'un champ lexical²³), (iii) soit la cooccurrence de deux ou plusieurs chaînes de caractères; – la constitution de telles collections peut être pratiquée notamment comme une étape d'exploration en même temps que de réduction de données initiales volumineuses; les résultats sont produits sous forme de fichiers qui peuvent à leur tour être traités selon le dispositif de base d'Anaconda tel qu'antérieurement exposé;

- est actuellement en développement un dispositif de modélisation/typologisation par comparaison et intégration automatique de graphes de synthèse tels que présentés plus haut et issus de sources multiples; – le propos correspond grosso modo à ceci : une série d'unités de données initiales, par exemple un certain nombre d'entretiens, livrent au terme de leur analyse autant de graphes de synthèse; un certain nombre de ces graphes peuvent présenter des homologies qui suggèrent leur participation à un « modèle » ou type commun qu'ils contribuent ensemble à manifester avec des degrés de consolidation variables notamment selon les échelles d'abstraction de la comparaison et le type de langage descriptif initialement produit par l'analyste; il s'agirait d'automatiser la production des résultats en la matière – donc de produire des « graphes de synthèse des graphes de synthèse » – à différents niveaux d'abstraction tout en rendant compte des degrés respectifs de consolidation, de manière notamment à créer des conditions où, stimulé à tenter des révisions de ses descriptions initiales, l'analyste pourrait aussi expérimenter les incidences de celles-ci sur les résultats de la modélisation; – on conçoit certes et par ailleurs que la modélisation à partir de graphes de synthèse multiples ouvre aussi à la possibilité de vérifier automatiquement la correspondance entre les modèles dégagés et les caractéristiques des sources et/ou acteurs qui supportent ces modèles, ceci notamment à l'échelle – et dans les limites – d'une sorte de micro-statistique²⁴... comme quoi toute chose, et donc aussi la fréquence, demeure là où elle a quelque pertinence.

Aspects techniques

Les procédures testées sous Anaconda sont en principe susceptibles de multiples réalisations aux détails variables et peuvent bénéficier d'outils de programmation divers. En pratique cependant, les développements actuellement disponibles ont été produits sous VBA – Visual Basic for Applications – le langage de programmation des applications de Microsoft Office : Word, Excel, PowerPoint... Puissant, souple et facile à utiliser²⁵, VBA

est accessible au très grand nombre : même si beaucoup l'ignorent, il fait normalement partie de l'équipement standard de tout détenteur de la suite bureautique citée. Si VBA permet d'automatiser ponctuellement des tâches en produisant des « macros »²⁶, il offre aussi le moyen de créer des logiciels de grande ampleur, tel notamment Anaconda.

Conçu pour programmer des applications préexistantes, VBA sert évidemment à développer des dispositifs qui usent des ressources de ces applications et qui, de ce fait, auront comme particularité d'être dépendants de celles-ci. Ainsi, Anaconda mobilise et exploite les potentialités du traitement de texte Word; aussi ne peut-il fonctionner sans celui-ci. Des points de vue militants pourraient le regretter, mais sans doute le feront-ils moins alors pour la dépendance elle-même que pour qui en bénéficie²⁷.

En contre-partie s'offrent notamment les avantages pratiques et techniques suivants :

- Anaconda se lance et s'affiche dans l'environnement même d'un traitement de texte largement répandu, comme s'il en était une fonctionnalité : il opère donc dans un cadre familier à un grand nombre d'utilisateurs;
- dans ce cadre, ces utilisateurs – hormis les fonctionnements contrôlés par Anaconda et qui obéissent à sa logique spécifique – gardent à leur disposition toutes les fonctionnalités du traitement de texte habituel ou qui sont accessibles à partir de celui-ci;
- pour la programmation, d'autre part, les fonctionnalités préexistantes ne sont évidemment d'aucune façon à produire à nouveaux frais ni distinctement : il suffit d'asservir de manière adéquate celles requises pour les propos spécifiques que les développements visent à servir;
- enfin, même si les praticiens des sciences sociales sont loin d'être tous des bricoleurs de soutiens logiciels, il n'est – pour la poignée de ceux qui le seraient tout de même, pour ceux aussi qui pourraient leur apporter un appui ou des contributions plus qualifiées, et finalement pour la production et le perfectionnement des soutiens logiciels eux-mêmes – il n'est sans doute pas sans intérêt de disposer en même temps d'outils de création, d'adaptation et de perfectionnement aussi accessibles et largement diffusés que ceux évoqués dans ce que nous clôturons ici.

Notes

¹ En zoologie, l'anaconda – *eunectes murinus*, ordre des Ophidiens, famille des Boïdés – est un boa aquatique du bassin amazonien, long de sept à onze mètres, non venimeux

et qui vit des fruits de son étreinte. Les deux derniers caractères flattent évidemment notre choix homonymique.

² A la différence du graphe qui suit, dès l'implication de trois disjonctions élémentaires, la représentation des structures croisées n'est techniquement plus réalisable qu'au moyen de matrices appropriées. Ces matrices peuvent également recevoir les données correspondant aux structures en éventail et parallèles; elles révèlent alors comment ces autres formes structurales, loin de former des natures sui generis, ne sont – à des échelles variables – que des réalisations réduites par rapport au maximum de combinaisons possibles qu'offre, pour un nombre donné de disjonctions élémentaires, l'articulation croisée de leurs termes respectifs; alternativement aux graphes, les matrices pourraient donc offrir une base commune aux descriptions et analyses comparées (Hiernaux, 1994).

³ Au simple sens étymologique : lié à l'organisation des lieux, sans compromettre ici ni une branche particulière des mathématiques, ni une dérive lacanienne.

⁴ Pour situer plus précisément les propositions qui suivent au sein des ressources logicielles destinées à l'analyse de corpus discursifs, voir Lejeune, 2007.

⁵ Allusion aux procédures typiques – et du reste tout-à-fait légitimes dans leur segment particulier – que soutient spécifiquement l'un ou l'autre logiciel dont la pénétration sur le marché se fait telle que la totalité des démarches non fréquentielles pourrait sembler lui appartenir.

⁶ Par exemple, lorsque après fermeture de fichiers articulés comme décrit, l'un d'entre eux est ré-ouvert pour reprendre le travail, ceux dont il dérive et/ou ceux qui dérivent de lui peuvent être automatiquement et simultanément ré-ouverts. Sous la tutelle du logiciel, il n'est par ailleurs pas possible de supprimer des fichiers participant à une série articulée sans être appelé à gérer les incidences sur celle-ci, etc.

⁷ Mutatis mutandis, des traitements analogues à ceux conçus pour les structures en parallèle pourraient s'appliquer aux structures en éventail ou croisées; considérablement moins fréquentes et moins amples ces derniers types de structures rendent cependant aussi moins urgents les soutiens logiciels à leur mesure.

⁸ Le propos, notons-le, n'est pas ici de valider l'analyse des données comme telle mais bien et uniquement d'illustrer la pratique d'indexation, à charge de l'analyste de justifier – dans et par les données – ce qu'il fait émerger de celles-ci.

⁹ Lors d'une toute première insertion, le sigle d'une polarité disjonctive – 'A ou 'B – est en principe indifférent. Par la suite, l'analyste tiendra compte des associations observées d'une sélection à l'autre.

¹⁰ Ainsi, en examinant anticipativement la Figure 7, espace/fichier 2, on comprendra que c'est seulement parce que l'on a réuni dans le plan des « états de soi » – 'SOI – tous les extraits repris en colonnes 1 et 3 que peut surgir la question de savoir si, à l'échelle des disjonctions structurant ce plan, il est plus optimal pour une finalité de recherche donnée de mettre en avant trois disjonctions – BESTIAL / NON-BESTIAL, INCROYANT / CROYANT et IRRÉGULÉ / RÉGULÉ – ou, à une échelle d'abstraction supérieure et valant pour l'ensemble, une seule disjonction que pourraient

reprendre les vocables IRRÉGULÉ / RÉGULÉ en admettant que la bête, l'incroyant et celui qui ne respecte rien ne seraient tous que des manifestations de l'être irrégulé.

¹¹ Pour l'illustrer, anticipons en examinant la Figure 7, espace 3, où le plan 'SOI est structuré par trois disjonctions, chacune correspondant à une des totalités décrites en colonne 3.

¹² Le moment où une collection atteint l'échelle requise pour que puisse en émerger l'esquisse de son principe, le moment de la « saturation », dépend tout autant de ce qui s'y donne à voir que du regard qu'y porte un analyste donné : certains verront vite, d'autres auront besoin de plus de temps et/ou de données, d'où – pour le soutien logiciel – la nécessité de laisser place à des conjonctures et rythmes variables.

¹³ Si l'illustration contient souvent la correspondance entre une catégorisation et un seul extrait, on se souviendra que l'espace réduit de cette illustration n'aura pas permis ici de montrer le cas le plus fréquent : la catégorisation unique pour une multiplicité d'extraits au contenu convergeant mais aux formes variables.

¹⁴ Ainsi la catégorisation BESTIAL qui figure deux fois en Figure 7, espace/fichier 2, colonne 2, n'est reportée qu'une seule fois dans le graphe de l'espace/fichier 3 et il en irait de même de toutes autres redondances, y compris nombreuses : le graphe de synthèse ne se complète donc que lorsqu'une catégorisation nouvelle est attribuée à au moins un extrait; dans les collections d'extraits, par ailleurs, l'association d'une même catégorisation à plusieurs extraits les identifie comme autant d'occurrences de l'unité de sens correspondante, occurrences qu'il est possible de rejoindre automatiquement en marche-arrière à partir de la catégorisation insérée au graphe de synthèse.

¹⁵ C'est le cas, en Figure 7, des catégorisations CROYANT et INCROYANT qui, dans l'espace/fichier 2 et bien que formant un couple, occupaient des lignes différentes du fait de la localisation initiale des extraits dont ils étaient issus. – Voilà pourquoi dans les collections d'extraits – et sans même considérer les difficultés qu'il y aurait à le faire, il n'importe guère de placer en face à face ceux qui se rapporteraient respectivement aux polarités 'A ou 'B d'une même totalité : cette organisation doit apparaître dans les résultats de l'analyse tandis que les étapes qui y conduisent peuvent et doivent disposer de la labilité qui leur convient.

¹⁶ Pour donner un seul exemple : la suppression dans l'espace/fichier des données à traiter d'une sélection qui soutiendrait, dans l'espace/fichier des collections d'extraits, l'unique extrait soutenant lui-même une catégorisation donnée dans l'espace/fichier du graphe de synthèse... modifierait directement ce dernier graphe lui-même en y supprimant la catégorisation concernée.

¹⁷ Incluant notamment que lorsque l'analyste insère pour la première fois la marque 'NNN'A, celle-ci est intégrée à la liste de même que 'NNN'B et que, lorsque la dernière sélection portant une marque d'un radical déterminé est supprimée ou ré-indexée, cette marque – sauf objection de l'analyste – est ôtée de la liste.

¹⁸ L'espace réduit des illustrations en Figure 7 plus haut, espace/fichier 2, n'aura pas permis d'y visualiser ces éléments. En grandeur nature, cependant, chaque extrait au sein d'une collection est – sur le modèle du référencement des paragraphes tel que présenté en Figure 7 plus haut, espace/fichier 1 – muni du nom du fichier et du numéro

du paragraphe de la sélection dont il procède ainsi que de la reproduction des marques d'indexation de cette sélection, lesquelles, lorsqu'elles sont multiples, livrent immédiatement l'information sur la localisation des « doublets ». – Ici comme ailleurs, Anaconda veille ainsi à ce que ses résultats demeurent lisibles même hors son cadre, par exemple sous forme de simple affichage de fichiers ou de consultation de copies imprimées.

¹⁹ Lorsqu'il a fait d'un terme la manière d'exprimer l'inverse d'un autre, l'analyste ne peut en effet modifier l'un quelconque des membres du couple sans notamment (i) modifier de la même façon toutes les occurrences de ce couple, (ii) avoir vérifié si les termes de remplacement n'engagent pas eux-mêmes d'autres couples et, le cas échéant, avoir géré la situation et ses incidences.

²⁰ On se souviendra de ce que les tableaux d'extraits, à la différence des graphes de synthèse, ne visent pas à placer en face à face sur une même ligne les éléments respectivement 'A et 'B correspondant à une même totalité. Lorsqu'un tel alignement survient néanmoins, comme en première et en dernière ligne du tableau de droite de notre illustration, il s'agit d'un effet purement aléatoire. En fait et comme exposé, le reclassement des extraits s'opère distinctement à l'échelle de la collection soit 'A soit 'B d'une section de tableau. Le tableau de droite de notre illustration présente un aboutissement où chacune de ces deux collections a été reclassée selon l'état des catégorisations attribuées à ses extraits.

²¹ Un seul exemple pour donner à comprendre : Anaconda reconnaît ses marques d'indexation au fait qu'elles commencent et finissent par le caractère °. Ce caractère ne peut donc figurer spontanément dans les données sans risquer de dérouter certaines procédures automatiques. Si c'est le cas, il doit donc être supprimé ou remplacé... ce qui, dans le cas de ° (code 176), se présente très favorablement : le caractère alternatif ° (code 186; oui, il est différent du précédent – regardez bien!) existe en effet et peut servir de remplacement à un grain près.

²² Ces informations serviront ultérieurement de base au référencement automatique des extraits dérivant des sélections opérées dans les données. Leur insertion aux paragraphes des données initiales tient cependant et une fois de plus au souci d'Anaconda de soutenir y compris les pratiques les plus « primitives » d'un quelconque analyste, en l'occurrence la copie et/ou le découpage des paragraphes avec des ciseaux ou encore la pratique du copier-coller avec l'ordinateur, deux variantes pour lesquelles il est bon de munir des extraits devenus autonomes de l'adresse de leur origine, adresse que l'on pourra également retrouver dans le fichier initial.

²³ On peut penser, par exemple, au champ familial avec les chaînes « père », « mère », « frère », « sœur », « oncle », etc.

²⁴ Pour illustrer la perspective de manière ne fut-ce que primitive, on conçoit comment, devant un modèle auquel participeraient des locuteurs des deux sexes, on pourrait vérifier le taux de contribution de chacun de ceux-ci, soit en soi, soit même relativement à la probabilité qu'en offrent les effectifs au départ.

²⁵ S'il n'était que la documentation qui l'accompagne d'office n'est pas des plus immédiatement pénétrables aux novices... et le dire ainsi est sans doute une litote.

²⁶ Pour « macro-instruction » : séquence d'instructions préenregistrées pouvant être répétées par l'ordinateur (www.le-dictionnaire.com).

²⁷ Comme le rappellent les noms de famille ci-dessous, il n'y aurait de ce point de vue aucune différence à dépendre de Microsoft Visual Basic plutôt que de Microsoft Word, même si le premier – VB, sans le « A » de « for Applications », un « grand frère » de VBA obéissant au même langage – produit des applications au fonctionnement totalement autonome.

Références

Greimas, A.J. (1966). *Sémantique structurale*. Paris : Larousse.

Hiernaux, J.P. (1978). *L'institution culturelle. Méthode de description structurale*. Louvain-la-Neuve : Presses Universitaires de Louvain.

Hiernaux, J.P. (1994). « *Et hic tres unum sunt* ». *Structures croisées et théorie des réductions*. [Note de travail inédite disponible auprès de l'auteur], 940430.

Hiernaux, J.P. (1995). Analyse structurale de contenus et modèles culturels. Application à des matériaux volumineux. Dans L. Albarello, F. Digneffe, J.P. Hiernaux, C. Maroy, D. Ruquoy, & P. de Saint-Georges (Éds), *Pratiques et méthodes de recherche en sciences sociales* (pp. 111-144). Paris : Armand Colin.

Lejeune, Ch. (2007). Petite histoire des ressources logicielles au service de la sociologie qualitative. Dans C. Brossaud, & B. Reber (Éds), *Humanités numériques, Tome 1. Nouvelles technologies cognitives et épistémologie* (pp. 197-214). Paris : Lavoisier.

Hiernaux, Jean-Pierre est docteur en sociologie, professeur au département des sciences politiques et sociales de l'Université Catholique de Louvain et membre du Laboratoire d'anthropologie prospective de la même université. Depuis la réalisation de sa thèse doctorale - L'institution culturelle : systématisation théorique et méthodologique (1977) - il n'a cessé de se consacrer aux développements propres à saisir dans différents contextes la composition et la dynamique des systèmes de sens socialement produits et socialement opérants. Il a contribué dans ce cadre à la mise au point de procédures d'analyse structurale de contenu dont il assure depuis de nombreuses années les formations en second et troisième cycles en même temps que des applications pratiques en divers domaines. Plus récemment, il s'est intéressé aux possibilités d'offrir à de telles analyses des supports logiciels appropriés, quitte à les produire soi-même.