

CAHIER DE ReCHERCHE

NOVEMBRE 1995



N° 82

ANALYSE TEXTUELLE
ET
STRUCTURES NARRATIVES DE RÉCITS

Valérie BEAUDOUIN

CRÉDOC
L'ENTREPRISE DE RECHERCHE

CREDOC

**Analyse textuelle
et
structures narratives de récits**

Valérie BEAUDOUIN

avec la collaboration de Claire EVANS

Département Prospective de la Consommation

NOVEMBRE 1995

142, rue du Chevaleret
7 5 0 1 3 - P A R I S

TABLE DES MATIÈRES

1. VISIONS DU MONDE ET ANALYSE DU DISCOURS.....	4
1.1. ANALYSE DES DONNÉES ET « VISIONS DU MONDE ».....	5
1.2. LA SÉMIOGRAPHIE ET L'ANALYSE DU DISCOURS	10
1.3. ANALYSE DE RÉCIT	12
2. PARCOURS D'INSERTION, TRAJECTOIRES DISCURSIVES	13
2.1. CONSTRUCTION DE PARCOURS À PARTIR DES RÉCITS	17
2.1.1. Repérer les étapes des parcours	17
2.1.2. Reconstruire les parcours.....	19
2.2. QUELQUES CONTRAINTES LIÉES À L'ANALYSE LEXICALE.....	21
2.2.1. Normalisation des sigles et des expressions composées.....	21
2.2.2. Réduction de la taille du vocabulaire.....	22
2.3. APPLICATION.....	25
2.3.1. Situations de départ et étapes de parcours	25
2.3.2. Flux de passage.....	30
2.3.3. Typologie des parcours.....	32
2.3.4. Quelques résultats complémentaires.....	34
3. CONCLUSION.....	38
4. BIBLIOGRAPHIE.....	39
5. TABLE DES ILLUSTRATIONS.....	42

L'année 1995 a été l'occasion de mettre au point une technique pour repérer la structure de récits. Gérard Genette (1972) distingue trois sens du terme *récit* :

- « *récit* désigne l'énoncé narratif, le discours oral ou écrit qui assume la relation d'un événement ou d'une série d'événements. » ;
- « *récit* désigne la succession d'événements, réels ou fictifs, qui font l'objet de ce discours, et leurs diverses relations d'enchaînement, d'opposition, de répétition, etc. ».
- « *récit* désigne encore un événement : non plus toutefois celui que l'on raconte, mais celui qui consiste en ce que quelqu'un raconte quelque chose : l'acte de narrer pris en lui-même. »

G. Genette propose trois désignations pour distinguer ces trois sens du terme *récit* : « je propose (...) de nommer *histoire* le signifié ou contenu narratif, *récit* proprement dit le signifiant, énoncé, discours ou texte narratif lui-même, et *narration* l'acte narratif producteur ».

Bien évidemment, ici l'objet d'étude sera principalement le récit tel qu'il vient d'être défini. Cela pour une raison assez simple liée à l'objet et au fonctionnement de nos métiers. Lors d'entretiens, de recueil de récits de vie, l'enquêteur n'a aucun accès à l'histoire réellement vécue : l'enquête n'est pas en général une auto analyse. Ensuite, ce ne sont pas toujours les mêmes personnes qui interrogent et analysent les textes, et l'analyse se fera sans doute en ignorant dans une large mesure l'acte narratif lui-même. Celui-ci suscite cependant un vif intérêt lorsque l'on étudie par exemple l'interaction entre enquêteur et enquêté. En règle générale, nous travaillons habituellement sur du discours oral retranscrit et c'est donc uniquement au récit, tel qu'il vient d'être défini que nous pouvons nous intéresser.

Nous avons été amenés à mettre au point une méthode pour analyser des récits sur un corpus de 580 parcours d'insertion de jeunes en difficulté. Les résultats se sont avérés particulièrement satisfaisants en raison de l'extrême homogénéité des récits. Cette même méthode a été appliquée à d'autres corpus avec malheureusement moins de succès. Ainsi avons-nous tenté de mettre à jour la structure narrative des tragédies, dans la mesure où, bien que la tradition distingue fortement genre narratif et genre dramatique (en raison de la forme dialogue), il existe bien un déroulement narratif dans la tragédie. Nous avons également fait des tentatives sur un ensemble d'entretiens sur le sida réalisés par le CES (Centre d'études sociologiques) de l'université Saint-Louis de Bruxelles. Ces entretiens cherchaient à appréhender l'attitude des individus par rapport au risque mais étaient en même temps l'occasion de « raconter leur vie ».

Les résultats ne sont pas satisfaisants et ne permettent pas de faire émerger des structures narratives solides. La raison de ces échecs relatifs est double :

- on disposait finalement d'un nombre trop faible d'entretiens et de trajectoires (moins de trente), ce qui est insuffisant par rapport aux 580 portraits de jeunes ;
- la structure narrative est beaucoup plus difficile à identifier dans les trajectoires où la narration est diluée dans le dialogue. De même, il est difficile de repérer dans des entretiens semi-directifs ce qui correspond à des récits de vie.

Il nous faudrait par exemple collecter des récits de vie nombreux auprès de populations homogènes, un corpus de déclarations d'incidents auprès de compagnies d'assurances pour tester avec plus de succès cette méthode sur d'autres corpus.

Nous ne désespérons pas de retrouver des corpus de récits aussi satisfaisants que les portraits de jeunes.

Cette méthode est née de la rencontre de deux logiciels ALCESTE et DISCAN (Discourse Analyser), plus précisément de la rencontre de deux méthodologies : la méthodologie ALCESTE, mise au point par Max Reinert et l'analyse de discours telle que l'envisage Pierre Maranda.

- La première consiste à identifier, dans les textes, divers mondes lexicaux ou champs lexico-sémantiques qui reflètent des « visions du monde » particulières. Ces mondes lexicaux sont obtenus à l'aide de procédures statistiques d'analyse des données qui rapprochent des fragments de texte employant le même type de vocabulaire. Chacun d'eux peut être investi à diverses reprises au cours du récit.

- La seconde méthodologie a pour objectif d'identifier les « éléments significatifs » (ce que l'on fait classiquement en analyse de contenu) à l'aide de procédures semi-manuelles, puis d'étudier les enchaînements de ces éléments à l'aide de procédures markoviennes. L'analyse du discours introduit une dynamique narrative dans l'analyse de contenu.

En bref, la combinaison des deux approches permet d'identifier les visions du monde, puis d'évaluer les itinéraires empruntés par les récits à travers ces univers. On est ici assez proche des méthodes de résumé automatique, la différence majeure étant qu'ici, les résumés ne peuvent être construits que si les récits sont très nombreux et redondants.

Bien que ce cahier de recherche ne parvienne pas encore à le démontrer, ce type de méthode devrait avoir de nombreuses applications dans nos domaines de recherche. L'analyse biographique, qui connaît un succès croissant en sciences sociales (Bauer et Maresca, 1992), devrait fort bien se prêter au type d'analyse que nous présentons. Il faudrait que les décalages entre le temps de l'histoire et celui du récit ne soient pas trop grands, en d'autres termes que le récit respecte la chronologie, pour que l'analyse des récits nous permette d'accéder aux histoires. L'application des méthodes d'analyse lexicale à l'analyse des entretiens semi-directifs s'est révélée jusqu'à présent assez décevante et, en tout cas, bien plus pauvre que ce qui peut être obtenu par une analyse de contenu fine. Cependant, quand le nombre d'entretiens dépasse la quarantaine et que l'analyse manuelle devient vraiment laborieuse, l'utilisation de notre méthode d'analyse de récit permettrait de dégager la structure des entretiens et les thèmes abordés pour faciliter l'analyse de contenu. Ainsi pourrait-on aisément retrouver toutes les parties de l'entretien qui concernent un thème donné.

1. VISIONS DU MONDE ET ANALYSE DU DISCOURS

Dans cette première partie, nous décrivons l'approche de M. Reinert en la situant par rapport au courant de l'analyse des données, puis la sémiométrie telle que l'entend P. Maranda. Ceci débouchera sur la description de la méthode d'analyse de récit qui combine les deux approches.

1.1. ANALYSE DES DONNÉES ET « VISIONS DU MONDE »

La méthodologie ALCESTE (Analyse des Lexèmes Cooccurents dans les Énoncés Simples d'un Texte) mise au point par Max Reinert (1990, 1993) s'inspire du courant de l'analyse des données de J. - P. Benzécri. Nous commencerons par faire un bref historique sur ce courant et sur ses fondements philosophiques, avant de décrire d'une manière plus détaillée la particularité de l'approche de M. Reinert, puis la manière dont nous l'interprétons par rapport à l'analyse du récit.

Les origines de l'analyse des données remontent au début du siècle. Les psychologues ont été les pionniers de l'exploration des données multidimensionnelles et de l'analyse factorielle. Pour une histoire précise des origines de la mesure en sciences sociales, on pourra se reporter aux travaux d'Olivier Martin sur l'histoire de la mesure (1995). Au début du siècle, Spearman, psychologue anglais, en analysant les liens entre les résultats scolaires et les aptitudes sensorielles d'élèves, démontre l'existence d'un **facteur** général d'aptitude ou d'intelligence qui sera désigné plus tard par la lettre *g*. Ultérieurement, non seulement un mais plusieurs facteurs seront recherchés à partir de données de plus en plus nombreuses. C'est là que se situent les origines de l'analyse factorielle.

L'analyse des correspondances, branche de l'analyse factorielle, trouve ses origines chez Fisher dans les années 40, puis Kendall et Stuart en 1961 mettent au point des méthodes d'analyse canonique des tables de contingence. Ceci leur permet de calculer les paramètres servant à tester l'hypothèse d'indépendance entre lignes et colonnes. L'école américaine de l'analyse des données (Multidimensional scaling), très orientée vers la psychologie, s'exprime principalement dans la revue « psychometrika » (Kruskal, Shepard...).

Benzécri s'intéresse à ces méthodes à partir de 1963. Il rencontre Shepard chez Bell aux Etats-Unis et s'initie donc aux théories et méthodes du « multidimensional scaling ». Il revient en France et professe en 1964 un cours à la faculté des sciences de Rennes où il expose ces méthodes. Son apport principal est d'avoir mis en évidence toutes les propriétés algébriques

de la méthode et montré son intérêt : tester l'indépendance des lignes et des colonnes, mais surtout décrire comment les données s'éloignent de cette hypothèse en représentant par des "proximités" les associations existantes entre les lignes et les colonnes (Diday et Lebart, 1977, p. 16). Benzécri n'a pas été l'initiateur de ces méthodes d'analyse des données multidimensionnelles mais a énormément contribué à leur diffusion en France : beaucoup de publications, de revues et de disciples, qui s'ils ont souvent rompu avec le maître, continuent à effectuer leurs recherches dans ce domaine. Pour l'analyse des données textuelles, on peut citer Ludovic Lebart, Pierre Pagès, Max Reinert, André Salem et bien d'autres encore, comme en témoigne le succès des Journées d'Analyse Statistique des Données Textuelles (JADT, 1991 ; Anastex 1993)

Revenons maintenant à Benzécri. En effet, il a mis au point et permis la diffusion de très nombreuses méthodes d'analyse, mais avant tout c'est une ambition théorique et une ambition philosophique qui l'animaient lorsqu'il s'est intéressé à ces méthodes.

Benzécri s'est intéressé aux méthodes d'analyse des données comme outil de recherche en linguistique et non pas en psychologie (discipline qui a été à l'origine des développements les plus nombreux) : "C'est principalement en vue de l'étude des langues que nous nous sommes engagés dans l'analyse factorielle des correspondances" (Benzécri, 1981). Il avait pour ambition théorique d'ouvrir les portes à une nouvelle linguistique à une époque dominée par la linguistique générative. Benzécri s'oppose à la thèse idéaliste de Chomsky qui dans les années 60 considère « qu'il ne peut exister de procédures systématiques pour déterminer la grammaire d'une langue, ou plus généralement les structures linguistiques, à partir d'un ensemble de données tel qu'un recueil de textes que les linguistes nomment corpus » (Benzécri, 1981). Contre cette thèse, Benzécri propose une méthode inductive d'analyse des données linguistiques "avec à l'horizon l'ambitieux étagement des recherches successives ne laissant rien dans l'ombre des formes, du sens et du style" (Benzécri, 1981). En ce sens, il est assez proche des objectifs des distributionalistes (Bloomfield et Harris) qui escomptaient construire les lois de la grammaire à partir d'un corpus d'énoncés.

"Nous proposons une méthode portant sur les problèmes fondamentaux qui intéressent un linguiste. Et cette méthode (...) effectuera une abstraction quantitative, en ce sens que partant de tableaux de données les plus divers, elle construira par le calcul, des quantités qui pourraient mesurer des entités nouvelles, situées à un niveau d'abstraction supérieur à celui des faits recensés d'abord."

(Benzécri, 1981, p. 4)

En identifiant les facteurs, il ne fait pas de doute qu'une opération d'abstraction a été réalisée. L'ordinateur ne donne pas de nom, ni de sens à ces entités qu'il a abstraites ; c'est au spécialiste d'apporter une interprétation.

L'**ambition philosophique** de Benzécri est de redonner toute sa valeur à la démarche inductive et de s'opposer par là même à l'idéalisme :

"Car nous condamnons que, de principes reçus à la légère, *l'idéalisme* prétende par une dialectique, fût-elle sous *l'emprise des mathématiques*, tirer des conclusions sûres ; puis à cette déduction *a priori* nous opposons *l'induction* qui a posteriori, des faits observés veut monter à ce qui les ordonne."

(Benzécri, 1968, p. 11)

Il critique les théories idéalistes qui posent l'existence d'un modèle et en vérifient approximativement la pertinence avec l'observation. Si cette démarche est encore vaguement acceptable pour la physique, elle ne l'est pas du tout pour l'économie. En effet, il n'existe pas de situation expérimentale suffisamment schématique pour satisfaire l'économiste. Il doute qu'il soit possible de réduire un objet complexe en un combinaison d'objets élémentaires, "car l'ordre du composé vaut plus que les propriétés élémentaires des composants" (Benzécri, 1968, p. 16).

L'objectif qu'il lui paraît possible d'atteindre avec l'analyse des données est de pouvoir dégager "de la gangue des données le pur diamant de la véridique nature". Et il fait référence au mythe de la caverne : à travers les ombres que sont les données, il faut remonter à des entités plus abstraites, aux idées. Cette démarche est rendue possible à ses yeux grâce à l'analyse des données et au « novius organum » qu'est l'ordinateur.

"Les moyens de calculs nouveaux permettent de confronter des descriptions complexes d'un grand nombre d'individus, pour aboutir à placer ceux-ci sur des cartes planes ou spatiales, image fidèle et accessible à l'intuition de la nébuleuse des données initiales."

(Benzécri, 1968, p. 21)

"Somme toute, on voit comment l'analyse affranchit des idées a priori. Des données aux résultats, l'ordinateur, insensible aux espérances comme aux préjugés du chercheur, procède sur la base ample et solide de faits définis et acceptés d'abord dans leur ensemble, puis dénombrés et ordonnés suivant un programme qui, parce qu'il ne sait pas comprendre, ne sait pas non plus mentir".

(Benzécri, 1968, p. 24)

"Enfin, parmi toutes les idées a priori, souvent contradictoires, que chaque problème suscite en si grand nombre, un choix opportun s'opère : bien plus, l'idée qui a posteriori, après

examen statistique des données, semble avoir été a priori fort naturelle ne se serait pas toujours présentée d'elle-même à l'esprit."

(Benzécri, 1968, p. 24)

"Auxiliaire de la synthèse, l'ordinateur est un outil mental : après l'organum d'Aristote et le Novum Organum conçu par Bacon, n'est-il pas le Novius Organum, « l'outil le plus nouveau » ? "

(Benzécri, 1968, p. 24)

Ces ambitions sont sans doute démesurées, ont été et sont largement critiquées mais elles ont le souffle des grands projets. L'analyse des données offre certes un ensemble de méthodes exploratoires efficaces, mais elle n'a pas atteint le statut de théorie scientifique.

Le rôle de Benzécri est fondamental pour une autre raison inattendue : il a fait entrer la créativité dans l'aridité des statistiques. Benzécri a en effet ouvert le champ de l'interprétation et donné droit au délire interprétatif.

Les préoccupations de Max Reinert constituent une orientation particulière par rapport aux travaux de Benzécri et de la plupart de ses disciples :

« Il s'agit, non pas de comparer les distributions statistiques des « mots » dans différents corpus, mais d'étudier la structure formelle de leurs cooccurrences dans les énoncés d'un corpus donné. »

(Reinert, 1993, p. 9)

Il considère un corpus comme une suite d'énoncés élémentaires produits par un sujet-énonciateur. Ainsi le texte est modélisé dans un tableau qui contient en ligne les énoncés, qui portent la marque du sujet énonciateur et en colonne les mots ou lexèmes, qui renvoient à des objets du monde (sans aucunement préjuger de la "réalité" de ces objets).

« Un énoncé traduit davantage un *point de vue* particulier plutôt qu'une représentation, le *point de vue* : impliquant en son centre l'existence d'un « sujet » dans une certaine modalité du faire ou de l'être. (...) Notre hypothèse consiste justement à considérer le vocabulaire d'un énoncé particulier comme une trace pertinente de ce *point de vue*. Il est à la fois la trace d'un lieu référentiel et d'une activité cohérente du sujet-énonciateur. Nous appelons *mondes lexicaux*, les traces les plus prégnantes de ces activités dans le lexique ».

(Reinert, 1993, p. 11)

« Un monde lexical, est donc à la fois la trace d'un lieu référentiel et l'indice d'une forme de cohérence liée à l'activité spécifique du sujet-énonciateur que l'on appellera *une logique locale*. »

(Reinert, 1993, p. 9)

Grâce aux procédures statistiques, qui rapprochent des énoncés employant le même type de lexique, on pourra identifier différents mondes lexicaux, qui pourront être révélateurs de différentes « visions du monde » ou de logiques. Ainsi dans l'étude sur les cauchemars d'enfants (Reinert, 1993), Reinert identifie trois types de logiques :

- une *logique de la séparation* où l'éloignement par rapport à la famille accroît les menaces extérieures,
- la *logique de la clôture*, où le sujet est plus passif, menacé dans son intégrité psychologique par l'agresseur
- et la *logique de la chute* : « le sujet n'existe plus, ni comme personne sociale ni comme personne physique puisque même sa pesanteur, qui l'ancre au sol, disparaît ».

Dans son étude sur Aurélia de Nerval (Reinert, 1990), il identifie trois types de mondes en classant les énoncés : le monde imaginaire, le monde réel et le monde symbolique qui portent chacun la marque d'un certain rapport au monde du narrateur.

Dans une autre perspective, la méthodologie ALCESTE est utilisée pour repérer les noyaux de la représentation sociale d'un objet (Lahlou, 1995).

D'un côté, on a une approche qui privilégie l'énonciateur —ou le narrateur si l'on se place dans la perspective du récit— et, dans l'autre, une approche qui privilégie la référence —ou l'histoire dans le cadre du récit—. Or pour l'analyse des récits, on cherche à identifier des classes d'énoncés narratifs, qui occupent une place intermédiaire : même si elles nous informent sur l'histoire et sur le narrateur, elles sont avant tout l'expression d'une structure de récit. Pour l'analyse de récits, on est donc amené à donner un autre type d'interprétation à la méthodologie ALCESTE.

1.2. LA SÉMIOGRAPHIE ET L'ANALYSE DU DISCOURS

La sémiographie est pour Pierre Maranda à la fois une méthode de terrain et un outillage pour l'analyse des données recueillies.

Aussi a-t-il mis au point un protocole de recueil de données, CAT-NAT (Continuous Association Tests and Narrative Association Tests (Maranda, 1992)), qui comprend des tests d'association libre et des tests d'association narrative. Trois stimuli sont proposés à la personne interrogée, qui, par écrit, fournit les associations libres pour chacun des termes et construit ensuite un récit faisant intervenir les trois stimuli, qui sont souvent des personnages, ou pour reprendre ses termes, des "dramatis personae".

Ainsi, Maranda et son équipe sont-ils parvenus à constituer un corpus écrit de plus de 6000 récits recueillis dans des zones aussi variées que le Canada, les îles Salomon, l'Amérique du Sud, la France... dans des régions développées comme sous-développées, urbaines comme rurales...(Maranda, 1992). Les trois "dramatis personae" étaient : l'homme, la femme et le serpent. Anthropologue de métier, Maranda s'intéresse essentiellement aux mythes. Il cherche à mettre en évidence les pré-construits culturels et l'incidence des mythes sur les représentations.

Pour l'analyse de ces données, Maranda a mis en place une méthode et conçu des outils informatiques regroupés dans le logiciel MicroMot/Discan, pour la mettre en oeuvre. La méthode d'analyse de Maranda ou *analyse de discours* se veut plus fine que l'analyse de contenu. En effet, alors que l'analyse de contenu consiste à repérer les « éléments significatifs » d'un texte, l'analyse de discours en « aborde les principes de constitution dynamique » :

Au lieu de se restreindre à un recensement des unités de sens comme le fait la simple analyse de contenu, elle en aborde les interrelations selon les séquences qui constituent les messages du "texte".

(Maranda et Nze-Nguema, 1994, p. 26)

Dans le cadre de la sémiographie telle que la pratique Pierre Maranda, les textes sont normalisés avec l'assistance du logiciel (Micromot/Discan) afin de dégager les éléments significatifs. On passe du niveau des vocables à celui des descripteurs sémantiques, du

langage à un métalangage. Ensuite, et c'est là qu'intervient l'analyse du discours, sont mesurées les relations séquentielles entre les différents X_ièmes. C'est à ce moment-là que Maranda fait intervenir des chaînes de Markov.

Au début du siècle, A.A. Markov formalisait ses modèles de « probabilités en chaîne », qu'on appelle aujourd'hui « chaînes de Markov ». En 1913 et en 1923, il proposait deux exemples d'application qui concernaient l'étude de la succession des voyelles et des consonnes dans des textes littéraires. La première étude portait sur les 20 000 premières lettres d'Eugène Onéguine (Markov, 1913) et la seconde sur les 100 000 premières lettres d'un roman d'Aksakov. Pour cette étude, il utilise les chaînes multiliées à trois événements : il calculait la probabilité d'avoir une voyelle sachant que les deux lettres précédentes sont des voyelles (Petruszewycz, 1983).

Revenons au corpus de récits construits autour des termes *homme*, *femme* et *serpent*. L'hypothèse qui animait cette démarche était que chez les occidentaux l'influence de la Genèse interviendrait dans la construction des récits, tandis que dans les populations mélanésiennes, les narrations n'auraient rien de commun avec ce mythe. La seconde partie de l'hypothèse a été très nettement vérifiée. En revanche, il s'est avéré que chez les occidentaux, la référence à la Genèse n'apparaît que chez une petite minorité (Maranda, 1992). Chez les occidentaux, la majorité des récits sont construits selon le stéréotype suivant : le serpent effraie ou mord la femme et l'homme l'éloigne ou le tue, ce qui lui vaut des faveurs de la femme. Maranda a également pu mettre en évidence des différences sociales à travers la construction des récits. Ainsi, les récits dans les classes les plus défavorisées sont plus violents : le serpent mord, plutôt que d'effrayer, et l'homme tue le serpent plutôt que de l'éloigner. Dans les catégories les plus défavorisées, la variante suivante est assez fréquente : le serpent tue de sa morsure l'homme et la femme.

Une étude réalisée pour une compagnie d'assurances livre des résultats passionnants. Malheureusement, elle n'est pas publiée. Je reconstitue le récit à partir de notes prises lors du séminaire animé par P. Maranda à l'EHESS en 1994. J'espère ne trahir ni les résultats, ni l'interprétation qu'en a faite P. Maranda.

Au Canada, comme dans beaucoup d'autres pays, les jeunes hommes au volant constituent une catégorie à risque pour les compagnies d'assurances. Ils ont beaucoup plus d'accidents de la route que les autres conducteurs. L'équipe de P. Maranda a interrogé deux échantillons de jeunes hommes, l'un ayant eu au moins trois accidents, l'autre n'en ayant eu aucun. Il est apparu que les associations libres autour du terme « femme » étaient très différentes. Les premiers glissaient très vite vers l'érotisme et le plaisir, tandis que pour les

seconds les associations glissaient vers la fécondité et la maternité. Ces associations étaient liées d'une manière assez nette avec la manière dont ils se représentaient leur situation familiale : les premiers se considéraient comme célibataires, les seconds se voyaient mariés avec des enfants. Ainsi, tout se passe comme si les représentations qu'ils se font de leur situation familiale induisait un type de comportement particulier au volant : ceux qui s'imaginent père de famille prennent moins de risque au volant. En tout cas, ces tests d'association mettent en évidence le caractère prédictif des associations libres pour l'étude des comportements.

1.3. ANALYSE DE RÉCIT

ALCESTE permet d'identifier des visions du monde à l'intérieur du corpus ou, dans le cadre du récit, les éléments de base de la narration. DISCAN intègre un système de normalisation ou de postcodage qui permet de passer des vocables aux concepts et d'identifier des éléments significatifs, puis de mesurer les transitions entre éléments significatifs. Il apparaît clairement que la combinaison de ces deux outils pourrait être d'une aide précieuse pour l'analyse de récits : identification de narratèmes (noyaux narratifs de base) et ensuite organisation de leurs enchaînements. Pour dire les choses autrement, on utilise ALCESTE à la place de la phase du postcodage, comme substitut à l'analyse de contenu, et DISCAN pour l'analyse du discours.

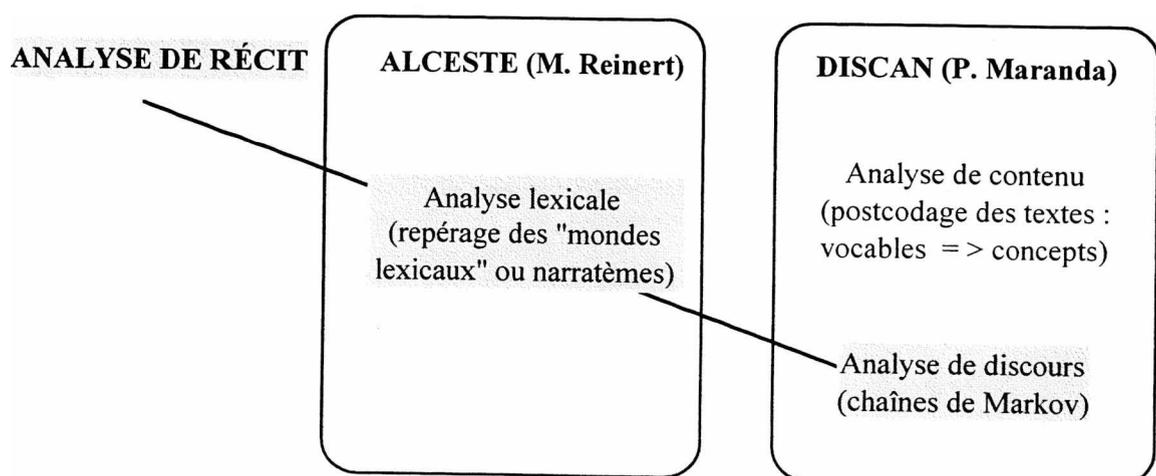


Figure 1 : Analyse de récit

Nous allons maintenant voir en détail l'adaptation de la méthode au cas particulier des récits de parcours d'insertion.

2. PARCOURS D'INSERTION, TRAJECTOIRES DISCURSIVES

Nous disposons d'un recueil de 580 récits de parcours de jeunes ("portraits de jeunes")¹ qu'il fallait analyser, structurer et résumer. La question posée était la suivante : comment analyser des récits de parcours trop nombreux pour être soumis à l'analyse qualitative traditionnelle et suffisamment homogènes pour faire l'objet de traitements quantitatifs ?²

Pour ce faire, nous avons été amenés à tester et adapter la méthode présentée précédemment. Ce choix de méthode nous permettait d'échapper à l'étape longue et fastidieuse du post-codage et, sans doute, de parvenir à une représentation plus synthétique des portraits.

Ces récits ont été rédigés par les conseillers des Missions Locales et PAIO³, et décrivent les trajectoires de jeunes qu'ils ont accompagnés dans leurs parcours d'insertion. Ces récits respectent un canevas précis : présentation rapide de la situation familiale, scolaire et sociale du jeune lors de son premier accueil à la Mission Locale, puis description des différentes étapes du parcours et enfin conclusion et bilan. La figure 2 présente un exemple de portrait.

Ce corpus de récits est entaché de nombreux biais.

Tout d'abord, il n'est pas représentatif de la population accueillie dans les Missions Locales et PAIO. Les portraits qui ont été choisis sont principalement des portraits *exemplaires* de trajectoires plutôt réussies. La DIJ a recueilli ces portraits à l'occasion du dixième anniversaire du réseau des Missions Locales et PAIO et souhaitait donner une image positive de l'activité des structures d'accueil.

D'autre part, les parcours ne sont pas toujours achevés dans la mesure où tous les jeunes n'ont pas en fin de portrait un emploi stable. Ils sont "censurés à droite" par la date de rédaction. Cependant tous les parcours, même s'ils ne sont pas tout à fait achevés, sont déjà bien avancés.

Enfin, et c'est sans doute le biais le plus important, les portraits ont été rédigés par les conseillers des Missions Locales et PAIO. Ceux-ci expriment dans ces textes leur propre point de vue sur l'insertion professionnelle et sociale. Les conseillers mobilisent l'ensemble du

¹ Ce travail a été réalisé à la demande de J. P. Bompard de la DIJ (Délégation Interministérielle à l'insertion des Jeunes).

² Cette partie présentée aux deuxièmes journées d'étude du CEREQ-LASMAS-IDL devrait faire l'objet d'une prochaine publication. Les résultats complets de l'étude figurent dans Beaudouin et Aucouturier (1994).

³ PAIO : Permanence d'accueil, d'information et d'orientation.

vocabulaire technique de l'insertion en conformité avec la représentation politique de l'insertion professionnelle. Il s'agit donc de parcours construits par une institution.

On dispose d'un matériau qui a l'avantage d'être homogène, mais qui, comme tout corpus issu d'une population spécifique, est fortement marqué par ses conditions de production. Notons que par-delà la prise en compte de ce biais, ceci signifie que la méthode mise au point sur ces discours bénéficie de cette stéréotypie caractéristique du genre "récit administratif" : il n'est pas sûr qu'elle puisse être appliquée à d'autres corpus de récits avec autant de succès.

Bien que non représentatifs et parfois inachevés, ces récits de parcours constituent un ensemble intéressant tant par la forme que par son contenu :

- ils forment un corpus cohérent et homogène qui fait appel à un champ lexical très spécifique (celui des dispositifs de l'insertion professionnelle). Il est donc particulièrement bien adapté aux techniques d'analyse textuelle ;
- ils donnent une vision globale de l'insertion qui ne se limite pas à la formation et à l'emploi : tous les aspects de l'insertion (sociaux, familiaux, psychologiques, sanitaires...) sont ici pris en compte et leur interaction clairement mise en évidence. Par ailleurs le rôle des échanges entre les jeunes et les conseillers occupe une place importante. Des représentations aussi complètes de l'insertion sont rares. En effet, la plupart des études quantitatives sur les trajectoires professionnelles se limitent souvent à la formation et à l'emploi (Ourtau et Werquin, 1994).

Une première partie sera axée sur la présentation de la méthode mise au point pour structurer les récits de parcours. Ensuite, on insistera sur les deux principales difficultés rencontrées pour mettre en oeuvre cette méthode sur le corpus de portraits : l'abondance des sigles et noms composés et la taille du vocabulaire. Enfin, les principaux résultats seront présentés, à titre démonstratif.

Mission Locale des Trois Vallées (91)

Laurent
24 ans

Laurent vit chez ses parents mais connaît de grandes difficultés relationnelles avec son frère. Il a abandonné sa scolarité en classe de 3e et a fait jusque-là, des petits boulots et travaux précaires. Son projet professionnel est orienté vers les métiers du spectacle après avoir travaillé dans divers petits lieux musicaux et théâtraux.

Mars 1990 - juin 1991.

Laurent est motivé et déterminé lorsqu'il se rend pour la première fois à la Mission Locale. il entre en formation qualifiante dans le spectacle et sort de formation en juin 91.

La Mission repère chez Laurent une situation de fragilité sociale qui semble s'aggraver en raison de l'absence de débouchés professionnels

correspondant à la formation acquise. Mais Laurent persiste dans son projet et la mission lui propose de poursuivre ses recherches d'emploi.

Avril 1992

Laurent disparaît et ne reprend contact qu'en avril 1992. Ses recherches ont été vaines et sa situation personnelle s'est dégradée : rupture avec le milieu familial, sans domicile fixe, problèmes de toxicomanie (héroïne) et alcoolisme.

Mai 1992

Après de nombreux entretiens à la Mission Locale Laurent rencontre une association spécialisée dans la toxicomanie des jeunes.

Il accepte de participer à un sevrage en juillet et une expérience voile en août (organisée par le Club de prévention et la Mission Locale).

Septembre 1992

Il lui est proposé un emploi de barman qui lui redonne confiance .

Octobre 1992

il passe les tests d'entrée aux Docks de France sur présentation de la Mission Locale.

...

A ce jour Laurent est embauché sous contrat à durée indéterminée par la même entreprise.

Il revit chez ses parents. Il reste suivi par la Mission Locale. La relation de confiance établie entre la Mission Locale, Laurent, sa famille, et une qualité de partenariat local adapté ont permis à ce jeune de s'insérer professionnellement.

Figure 1 : Exemple de portrait

2.1. CONSTRUCTION DE PARCOURS À PARTIR DES RÉCITS

Qu'est-ce qu'un parcours d'insertion ? C'est le passage du jeune par une succession d'états, entre le moment où il entre en contact avec la Mission Locale et celui où il disparaît de son champ d'action. Pour nous, ce trajet dans la vie réelle n'est perçu qu'à travers un récit. Ces récits sont des trajectoires discursives, images textuelles des trajectoires réelles. Nous allons donc chercher à repérer les lieux de discours (description d'un passage en formation, de la recherche d'un emploi, de difficultés financières ou sociales...) qui correspondent aux lieux du réel que parcourt le jeune. A partir des récits, nous cherchons à extraire une carte, image textuelle du territoire de l'insertion. Et sur cette carte, nous replacerons les itinéraires du jeune.

La méthode reconstitue les parcours en deux temps : l'un analytique et l'autre synthétique.

- Dans un premier temps, on identifie les différentes étapes possibles des parcours, en faisant abstraction de la manière dont elles sont ordonnées. Pour cela, on utilisera les méthodes d'analyse statistique des données textuelles, en l'occurrence ALCESTE.
- Dans un deuxième temps, on résume chaque récit par l'enchaînement des étapes qui ont effectivement constitué le parcours. Cette reconstruction fait appel à des méthodes plus classiques comme les chaînes de Markov (ici DISCAN) et l'analyse séquentielle (classification sur des variables de transitions ou couples d'états).

2.1.1. Repérer les étapes des parcours

Les portraits sont en général découpés en paragraphes précédés par une date ou une période. Ces paragraphes sont censés refléter chacun une étape de parcours. Il suffit donc pour étudier les étapes possibles de classer ensemble les paragraphes "similaires", en faisant l'hypothèse que, s'il existe effectivement un territoire du réel auquel renvoient ces différents récits, les paragraphes doivent bien tous renvoyer à des lieux identiques de ce territoire, même s'ils sont décrits de manière légèrement différente dans chaque récit. Quel principe de similarité choisir ? On a opté pour un principe de proximité lexicale faute de pouvoir travailler avec des méthodes automatiques sur la proximité sémantique. Les méthodes d'analyse statistique des données textuelles (Benzécri et col, 1981 ; Lebart et Salem, 1994)

s'avèrent, pour le moment, les mieux adaptées pour identifier des « noyaux de sens » ou unités de base constitutives de la représentation (Lahlou, 1995) exprimées dans les textes. C'est principalement le logiciel ALCESTE mis au point par Max Reinert (1993) qui sera utilisé pour cette étape.

On constitue un tableau de présence-absence sur lequel sera appliquée une méthode de classification descendante. Ce tableau contient en ligne les portraits et autant de colonnes que de mots utilisés dans les portraits. A l'intersection d'une ligne i et d'une colonne j , on obtiendra la valeur 1 si le mot j est présent dans le paragraphe i et la valeur 0 sinon. On aboutit à une typologie des paragraphes qui pourra aisément être assimilée à une typologie d'étapes de parcours.

Ce tableau est soumis à deux classifications descendantes avec des paramètres légèrement différents. Les deux typologies sont comparées et l'on conserve les classes les plus stables. Tous les paragraphes appartenant à une même classe partagent le même univers lexical. Chaque classe est décrite par les mots les plus typiques qui lui sont associés et par les paragraphes qui contiennent le plus de mots typiques.

Cette analyse conduit, sur notre corpus, à une typologie en huit classes. L'examen de cette typologie montre que les deux premières classes correspondent à deux types de situations initiales et les six autres à différentes étapes de parcours. Ce sont des étapes abstraites dans la mesure où l'on ne sait pas, à ce stade, comment ces étapes s'ordonnent et s'enchaînent.

La figure 3 résume cette première étape analytique de l'analyse.

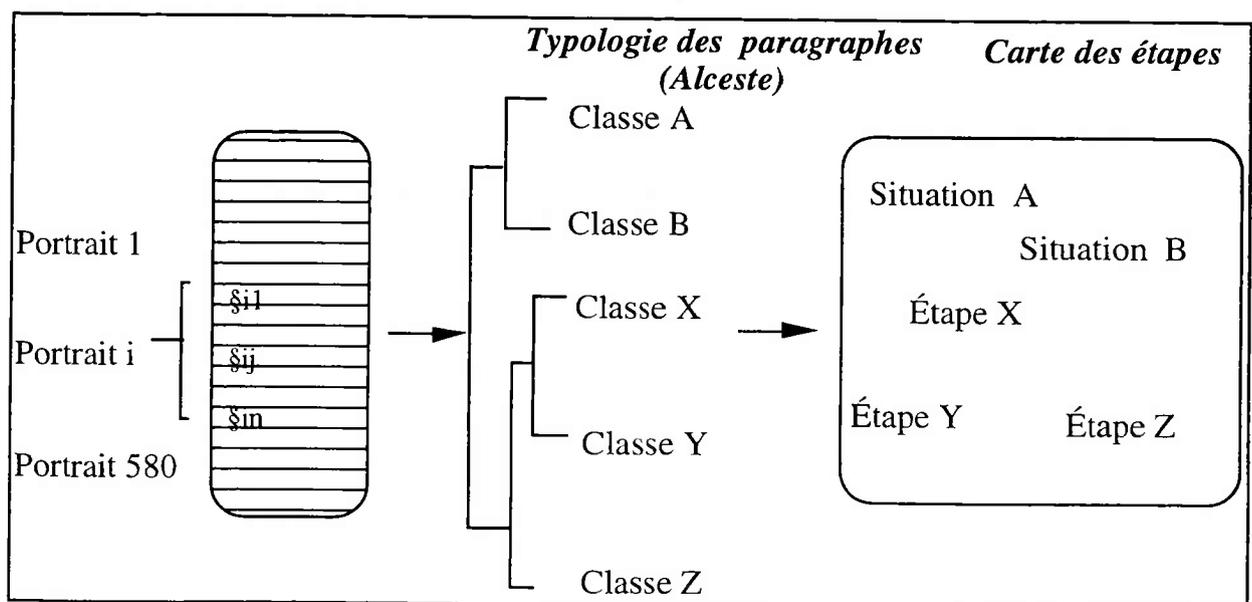


Figure 2 : Identification des étapes possibles

2.1.2. Reconstruire les parcours

Suite à la première analyse, chaque paragraphe des portraits appartient à une classe particulière qui correspond soit à une situation de départ, soit à une étape. Chaque portrait peut donc être résumé par la succession des étapes traversées que nous appellerons la "phrase-trajectoire". Par exemple, le portrait d'un jeune peut être résumé par la séquence :

Situation B - Étape X - Étape Z - Étape Y

On peut alors mesurer, sur l'ensemble des portraits, les flux de passage d'une étape à une autre. Pour la mesure des flux de transition, nous avons utilisé des chaînes de Markov à un degré (logiciel Discan de P. Maranda (1978, 1989)). On se situe alors en chacune des étapes et l'on mesure vers où s'orientent les récits étant passés par cette étape. Cette approche permet d'identifier les lieux de passage les plus fréquentés. Elle est cependant partielle puisqu'elle ne mesure que les passages d'une étape à la suivante et ne donne donc pas une image globale de la trajectoire.

Cette première approche a donc été complétée par une typologie des parcours construite sur la base des séquences de deux étapes (ou variables de transition). On retrouve "l'analyse de couples" présentée à plusieurs reprises aux Secondes Journées d'Analyse des Données Textuelles (1993). Il s'agissait en général d'exemples d'analyse par couples sur des étapes pré-codées au moment de l'enquête. Dans l'étude sur les portraits, cette analyse est faite sur des étapes identifiées par analyse lexicale. En bref, la typologie est obtenue en analysant un tableau qui contient en ligne les portraits et en colonne les couples d'étapes. Ainsi, le portrait précédemment résumé est décrit par les trois variables de transition suivantes :

Situation B - Étape X ; Étape X - Étape Z ; Étape Z - Étape Y

Alors que la première approche rend compte de l'extrême variété des parcours possibles, la seconde approche tend à caricaturer les types de trajectoires¹.

En fin de compte, comme le montre la figure 4, on obtient une carte des trajectoires avec les flux de passage et une typologie des parcours.

¹ Même si plus de 20 % des parcours sont irréductibles à un de ces quelconques types.

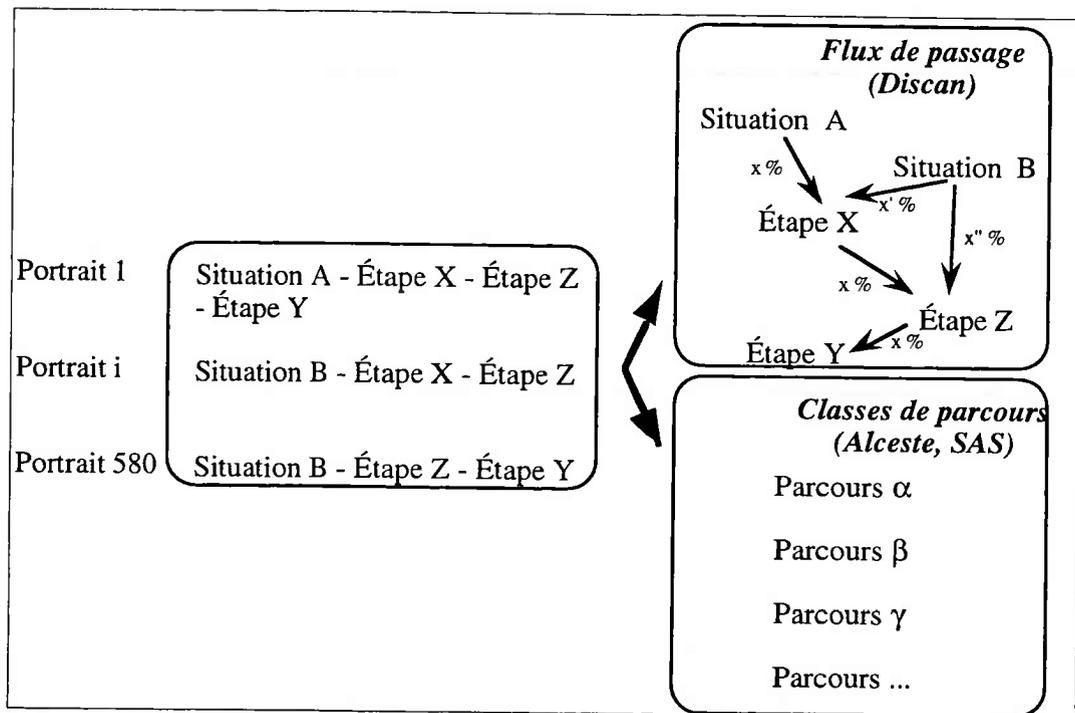


Figure 3 : Flux de passage et types de parcours

2.2. QUELQUES CONTRAINTES LIÉES À L'ANALYSE LEXICALE

Pour la première étape de l'analyse (identification des étapes possibles), il a fallu faire face à certaines contraintes liées aux particularités du corpus analysé et plus généralement à toute procédure d'analyse lexicale. Une longue étape de préparation du corpus a été indispensable, qui a tout d'abord consisté à identifier et normaliser tous les sigles et expressions composées, et ensuite à réduire par différentes opérations la taille du vocabulaire.

2.2.1. Normalisation des sigles et des expressions composées

La caractéristique stylistique la plus frappante de ces portraits est l'utilisation massive de sigles (*CFI, CES, PAIO, CAP, ...*) et d'expressions composées (*Mission Locale, contrat de qualification, contrat d'apprentissage, remise à niveau...*). C'est le propre des vocabulaires spécialisés d'être amenés à utiliser sigles et expressions composées pour désigner de nouveaux "objets" : ici de nouvelles mesures ou de nouveaux dispositifs d'aide à l'insertion. Le recours à ce vocabulaire très spécialisé reflète l'effort de conformité des conseillers par rapport aux divers outils et mesures d'aide à l'insertion. Pour que les portraits soient quand même lisibles par des non spécialistes, tous les sigles sont à la fois donnés sous leur forme développée et sous forme de sigle.

Dans les méthodes d'analyse des données textuelles, l'unité de base est la forme graphique, c'est-à-dire une "suite de caractères comprise entre deux caractères délimiteurs" (Lebart et Salem, 1988). Les caractères délimiteurs sont l'espace et tous les signes de ponctuation. Il a donc fallu faire en sorte que chaque sigle et chaque expression composée soient considérés comme une seule unité d'analyse, comme un seul "mot". Cette démarche se justifie dans la mesure où un sigle ou une expression ne représentent qu'un unique "objet" : une mesure, une institution, un type de formation ... Elle se justifie également pour des raisons statistiques : deux mots systématiquement associés, s'ils ne sont pas considérés comme une seule unité, créent des distorsions dans la classification (Beaudouin et Lahlou, 1993).

Ainsi, toutes les formes développées des sigles et toutes leurs variantes ont dû être ramenées à la forme condensée du sigle. Les formes développées du sigle : *Association formation professionnelle pour adultes, Association pour la formation professionnelle des*

adultes, *Association pour la formation professionnelle*, *Association de formation professionnelle pour adultes* ont, par exemple, toutes été ramenées au sigle AFPA. Les expressions composées ont, quant à elles, été introduites dans un dictionnaire de locutions afin d'être considérées comme une seule unité ou une seule forme.

2.2.2. Réduction de la taille du vocabulaire

Une des difficultés classiques de l'analyse statistique des données textuelles est liée à la forme très particulière de la distribution des fréquences dans un texte (Loi de Zipf¹). Ainsi, on a toujours peu de mots très fréquents et beaucoup de mots très rares. Ces derniers sont évidemment peu pertinents pour l'analyse statistique, mais il est cependant dommage de les perdre. Afin que l'analyse soit la plus riche possible et que le minimum d'information soit perdu, on procède à des regroupements lexicaux pour que les formes rares ne soient pas d'emblée éliminées par les traitements statistiques.

Dans les portraits, les mots de fréquence inférieure ou égale à trois, peu aptes à jouer un rôle dans les traitements statistiques, représentent 63 % du vocabulaire. Bien qu'ils ne représentent que 4 % des occurrences, on a cherché à les conserver dans les analyses. Pour cela, différentes techniques ont été utilisées pour réduire la proportion du vocabulaire représentée par les basses fréquences et plus généralement la taille du vocabulaire.

Ces opérations ont essentiellement consisté à :

- lemmatiser, c'est-à-dire ramener les formes fléchies à leur forme canonique (formes conjuguées des verbes ramenées à l'infinitif, pluriel transformé en singulier...). Pour cette opération, on a utilisé le lemmatiseur SIT mis au point par Langage Naturel S. A. selon nos spécifications (CRÉDOC). Ce lemmatiseur s'appuie sur un vaste lexique constitué par F. Pigamo (1990) et sur l'analyseur syntaxique de P. Constant (1991). Grâce à la lemmatisation, les mots de fréquences 1 à 3 qui représentaient 63 % du vocabulaire n'en représentent plus que 37 %.

¹ ZIPF G. K., (1974 (1ère édition 1936)).- *La psychobiologie du langage : une introduction à la philologie dynamique*, RETZ-CEPL, Paris. Cf aussi Beaudouin, "Structure lexicale, registres et thèmes", Cahier de recherche n°61, CRÉDOC, 1994.

- stemmatiser, c'est-à-dire ramener à une même racine les mots qui en dérivent. Cette opération est effectuée par le logiciel ALCESTE (Reinert, 1983). Par exemple, les mots *confiant*, *confier*, *confiance* seront ramenés à la racine *confi+*.
- exclure de l'analyse une partie du vocabulaire (les mots grammaticaux —pronoms, prépositions, conjonctions, certains adverbes...— et les noms propres). Les mots grammaticaux ne sont pas pertinents pour notre propos puisqu'ils ne renvoient pas à des "objets du monde" ; ils jouent le rôle de "ciment" syntaxique. Le logiciel Alceste intègre des dictionnaires informatisés de mots grammaticaux : il est alors aisé de les exclure de l'analyse. Par ailleurs, on a constitué un dictionnaire¹ des noms propres présents dans les portraits, pour les exclure des analyses. Il est inutile en effet que les noms de lieux et de personnes interviennent dans les analyses portant sur le texte lui-même. Le prénom est en effet très régulièrement répété tout au long des portraits, ce qui aurait pu rapprocher artificiellement les portraits de jeunes ayant les mêmes prénoms (qui sont d'ailleurs factices pour maintenir l'anonymat).
- effectuer à la main certains regroupements.

Il nous a semblé nécessaire de regrouper les différentes variantes utilisées pour désigner les mesures d'insertion. En effet, un simple examen des *contextes*², nous a permis de mettre en évidence l'équivalence entre certaines expressions comme *entrer en qualification* et *entrer en stage de qualification*. Les regroupements que nous avons effectués sont les suivants :

¹ A l'aide d'un module du logiciel Hyperbase (développé par Étienne Brunet (1992 et 1993) CNRS-INALF) qui permet la reconnaissance des noms propres.

² C'est également un module d'Hyperbase qui nous a permis d'extraire les contextes de certains mots : *stage*, *qualification*, ...

Dans le texte	Uniformisé
stage d'alphabétisation alphabétisation	alphabétisation
stage de mobilisation mobilisation sur projet action de mobilisation stage de préparation à l'emploi	mobilisation
mise à niveau remise à niveau stage de remise à niveau	remise à niveau
stage de préqualification stage préqualifiant préqualification	préqualification
stage qualifiant stage de qualification action qualifiante action de qualification qualification	qualification
mesure exo-jeunes contrat exo-jeunes exo-jeunes	exo-jeunes

Figure 4 : Regroupements effectués à la main

Les choix de réduction qui ont été ici adoptés ont pour effet de renforcer la cohérence des portraits, mais naturellement, cela se fait au détriment de leur richesse.

2.3. APPLICATION

Nous présentons les différentes étapes de l'analyse : l'identification des étapes de parcours, la mesure des flux de passage d'une étape à l'autre et enfin la typologie des parcours¹. Nous explicitons la méthode utilisée et les résultats qu'elle permet d'obtenir.

2.3.1. Situations de départ et étapes de parcours

L'analyse lexicale des paragraphes permet d'identifier trois pôles (un pôle de situation et deux pôles d'interaction). Chaque pôle se divise en classes comme le montrent les deux graphiques suivants.

Dans la partie grisée des graphiques, se trouve la liste des mots caractéristiques de chacune des classes qui a permis d'interpréter les classes. Le *pôle de situation* correspond grosso modo aux paragraphes de cadrage qui ouvrent les portraits. Les deux autres pôles correspondent aux différentes étapes de l'interaction entre le jeune et le conseiller.

2.3.1.1. Situations de départ

Le premier pôle isole deux classes de paragraphes dont les mots caractéristiques servent à décrire une situation, un état : niveau scolaire, situation familiale, professionnelle ... Ces paragraphes correspondent à la description de la situation de départ des jeunes accueillis et de leurs difficultés : faible niveau scolaire, milieu défavorisé, conflits familiaux ou sociaux, problèmes de santé, de logement...

L'analyse lexicale des paragraphes permet de distinguer deux types de situations, grâce au type de vocabulaire employé. Les paragraphes de la première classe insistent sur le faible niveau scolaire et décrivent la situation familiale (le jeune, célibataire, vit souvent chez ses parents). Cette classe ("handicap") correspond aux jeunes qui ont principalement un besoin d'insertion professionnelle, du fait d'un **faible niveau scolaire** initial et de leur appartenance à un milieu défavorisé. Ils sont dans une situation de handicap social.

Les paragraphes de la seconde classe décrivent aussi une situation scolaire et professionnelle difficile (peu ou pas de formation, recherche d'emploi), mais surtout une situation de rupture familiale (conflit, séparation) ou sociale (délinquance, handicaps...). Les

¹ Pour une description plus détaillée, on pourra se reporter à Beaudouin et Aucouturier (1994).

jeunes de la seconde catégorie ("traumatisme") recherchent une insertion professionnelle, mais sont en plus dans une situation de traumatisme social ou psychologique liée à une **rupture familiale ou sociale**.

parent, vit, mère, célibataire, père, quitter, classe, soeur, scolarité, enfant, frère, niveau_vi, système scolaire, CPPN, famille, vivre., origine, habiter, p_sortir, collège, élever, niveau_v, p_marier, école, divorce, premier, SES, unique, ouvrier, grands-parents, retraité, issu, arriver, seul, défavoriser, p_scolariser, fratrie, préapprentissage, scolariser, invalide, décéder, accueil, titulaire.	Faible niveau scolaire, milieu défavorisé, (10 % des paragraphes, 54 % des portraits)	Pôle de situation
père, vivre., famille, mère, enfant, parent, aine, séparer, soeur, frère, âge, marier, appartement, quitter, foyer, ami, enfance, familial, fils, arriver, grave, famille d'accueil, vit, rupture, alcool, seul, divorce, p_placer, modeste, autonomie, violent, nombreux, perdre., couple, ressource, naître, retrouver, franc+, drogue, difficile, p_séparer, ASE, loger, RMI, mal, habiter, connaître., chômage, concubin, problème, provisoire, jeune femme, FJT, unir, beau, agriculture, délinquant, PJJ, tante, venir., national, juge, compagne, revenu, origine, maison, rejet, séjour, p_héberger, HLM.	Faible niveau scolaire + rupture familiale ou sociale (6 % des paragraphes, 30 % des portraits)	

Figure 5 : Situations de départ¹

2.3.1.2. Le suivi des jeunes : les étapes de l'insertion

Les deux pôles suivants regroupent les différentes classes de paragraphes qui correspondent aux étapes de parcours suivantes : *définition du projet professionnel, remise à niveau, aide financière, aide à l'insertion sociale, formation et emploi*.

Le premier pôle, celui de *l'interaction*, regroupe les quatre premières étapes. Il fait explicitement apparaître l'étroite collaboration entre les jeunes et les conseillers des structures d'accueil dans ce cheminement vers l'insertion. Le principe de déroulement de ces étapes est le suivant : la structure d'accueil fait une ou des propositions, le jeune fait son choix, s'investit et agit. La démarche adoptée prépare le jeune à accéder à l'autonomie.

Dans le second pôle, celui de *l'autonomie*, qui regroupe les deux dernières classes, le rôle de la Mission Locale ou PAIO devient secondaire : ce sont les étapes qui correspondent à l'accès du jeune à l'autonomie : obtention d'un diplôme, d'une formation en alternance ou d'un emploi.

¹ p_ marque les formes passives des verbes, et n_ les formes négatives.

Le graphique ci-dessous résume l'organisation des différentes étapes en fonction des deux pôles.

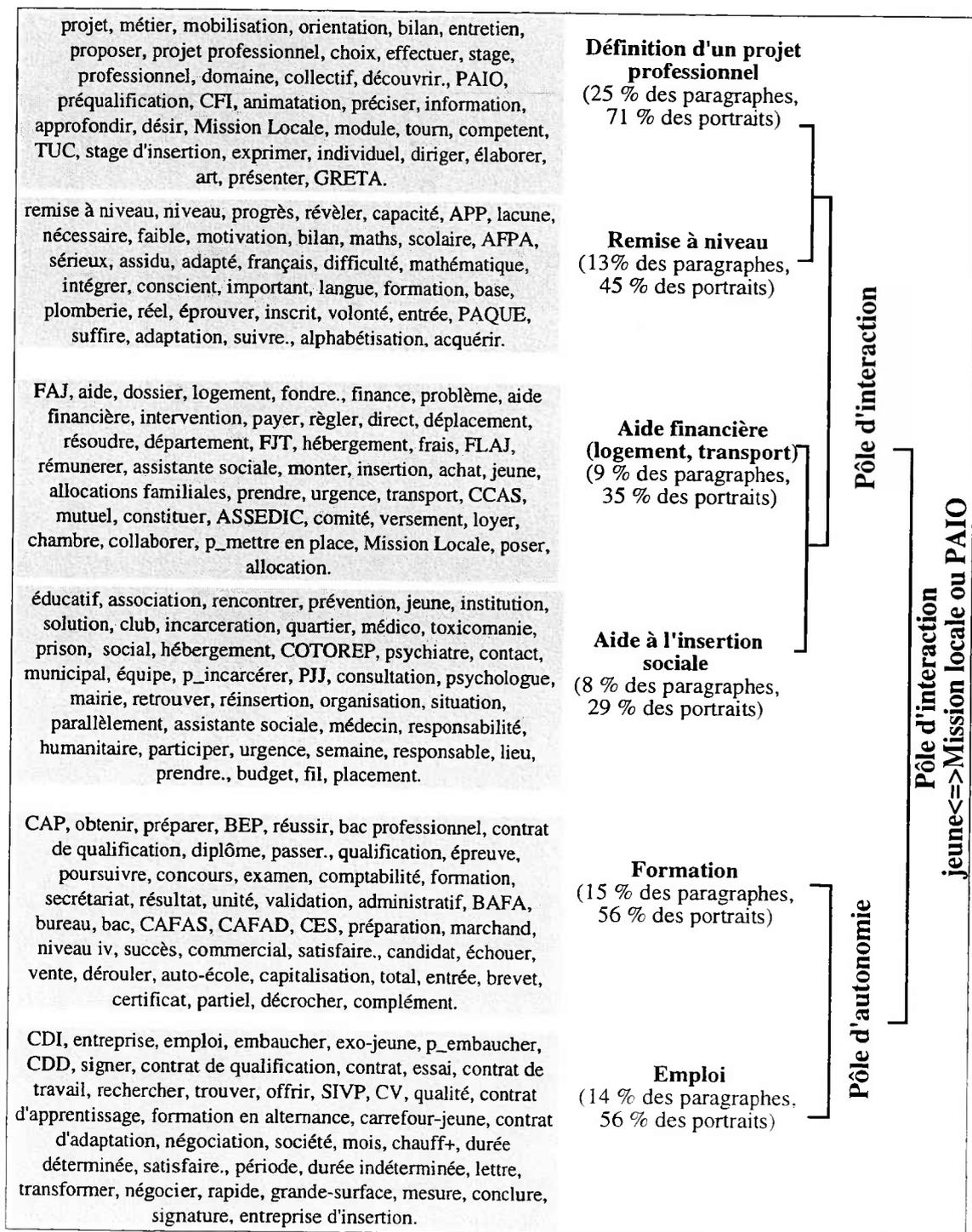


Figure 6 : Étapes de parcours

Parmi les quatre étapes de **l'interaction**, deux correspondent au coeur de l'activité des structures d'accueil, à savoir l'insertion professionnelle, et mobilisent un grand nombre de mesures publiques (stages, actions...) :

- la **définition d'un projet professionnel** : cette étape est présente dans 71 % des portraits. C'est indéniablement sur cette étape que les conseillers concentrent leurs efforts, en donnant des informations, en faisant des propositions, en bref en élargissant le champ des possibles pour les jeunes. Le rôle du dialogue est ici fondamental. Cette étape est aussi celle qui fait le plus appel à des mesures publiques : pour confirmer ou infirmer ses choix, le jeune passe par des "stages de découverte des métiers" et des "stages de mobilisation sur projet".
- la **remise à niveau** : dans 45 % des portraits, les jeunes qui ont un niveau scolaire trop faible passent par cette étape avant d'accéder à une formation professionnelle. Elle s'effectue à travers des stages (d'alphabétisation, de remise à niveau), des Ateliers Pédagogiques Personnalisés et souvent dans le cadre de la Préparation Active à la Qualification et à l'Emploi (PAQUE).

Les deux autres étapes de l'interaction sont davantage centrées sur l'insertion sociale. Elles concernent les difficultés financières et sociales rencontrées par les jeunes et les solutions qui sont proposées par les structures d'accueil. Au cours de ces étapes, comme dans les précédentes, le lien entre le jeune et la structure d'accueil est très étroit. Il s'agit de :

- **l'aide financière** : dans 35 % des portraits, les jeunes ont rencontré au cours de leur parcours des difficultés financières, principalement liées à l'hébergement ou au transport. Ils ont eu besoin d'argent pour payer un loyer ou un permis de conduire, pour acheter un moyen de locomotion ou un titre de transport... Certains ont ce type de difficultés dès qu'ils se présentent à la Mission Locale ou PAIO. Mais généralement, ils ont besoin de ressources financières supplémentaires au moment de l'accès à une formation ou à un emploi qui implique souvent une mobilité géographique. La Mission Locale les aide à obtenir un soutien financier auprès du Fonds d'Aide aux Jeunes (FAJ) ou d'autres organismes.
- **l'aide à l'insertion sociale** : 29 % des jeunes ont rencontré des difficultés d'insertion sociale, dont les symptômes n'apparaissent pas forcément lors du premier accueil : prison, toxicomanie, fragilité psychologique, handicaps physiques ou mentaux... Une fois que le problème est identifié, le rôle de la Mission Locale est de mobiliser le réseau des

partenaires locaux afin de mettre en place un suivi social actif pour lutter contre l'exclusion : associations, éducateurs, médecins, psychiatres...

Les deux étapes ci-dessus sont très fréquentes dans les portraits provenant des Missions Locales, mais n'apparaissent que rarement dans ceux des PAIO, dont l'action reste principalement centrée sur la formation et l'emploi. Depuis leur création, le champ d'intervention des Missions Locales est en effet beaucoup plus large que celui des PAIO, ainsi que l'éventail des mesures qu'elles peuvent mobiliser.

Les deux dernières étapes correspondent à l'accès à **l'autonomie**.

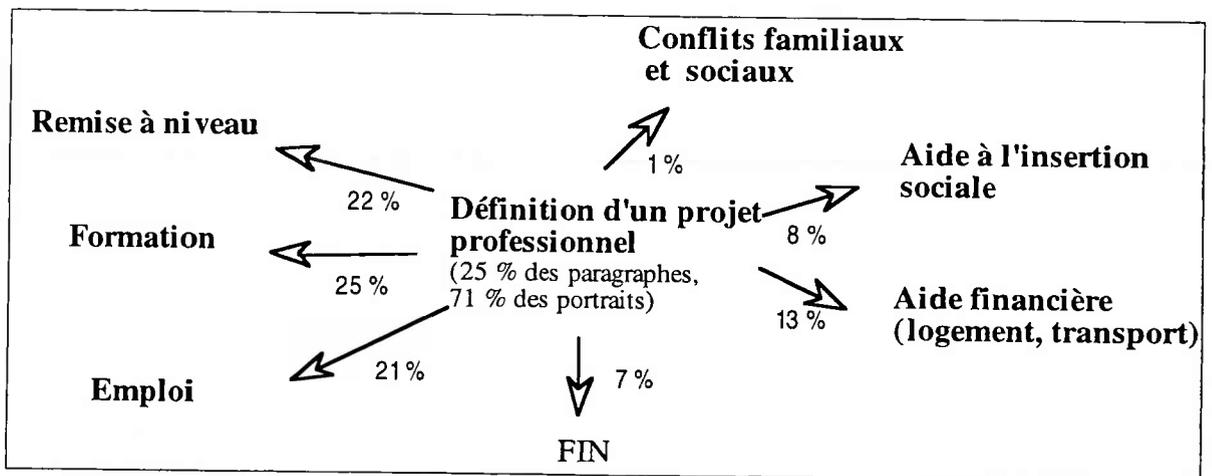
- **La formation**, qu'elle soit en alternance ou non, est la première étape vers l'autonomie. Dans ce cadre, le contrat de qualification est davantage considéré comme un moyen d'accès au diplôme que comme un contrat de travail. Dans les portraits, la notion de formation est fortement valorisée : elle est un moyen de redonner confiance au jeune, et de le placer dans une perspective de réussite. Le diplôme est considéré par les conseillers comme la garantie d'une meilleure insertion professionnelle. Dans 56 % des portraits, les jeunes passent par cette étape.
- La deuxième étape correspond à **l'emploi**, avec le plus souvent la signature d'un Contrat à Durée Indéterminée ou à Durée Déterminée (CDI ou CDD). Les contrats en alternance (contrats de qualification, d'apprentissage, d'adaptation) sont également très fréquents. Ils sont ici principalement considérés comme des contrats de travail et accessoirement comme des outils de formation. Cette étape est présente dans 56 % des portraits.

La présence de la formation en alternance dans ces deux étapes montre son ambiguïté fondamentale : tantôt elle est considérée comme un moyen de formation professionnelle, tantôt comme une forme d'emploi.

Viennent d'être identifiés deux types de situation de départ et six étapes de parcours. On obtient ainsi la représentation que les conseillers se font des étapes d'un parcours d'insertion et de leur rôle à chacun de ces stades. Il reste à voir comment s'ordonnent ces diverses étapes pour construire un parcours.

2.3.2. Flux de passage

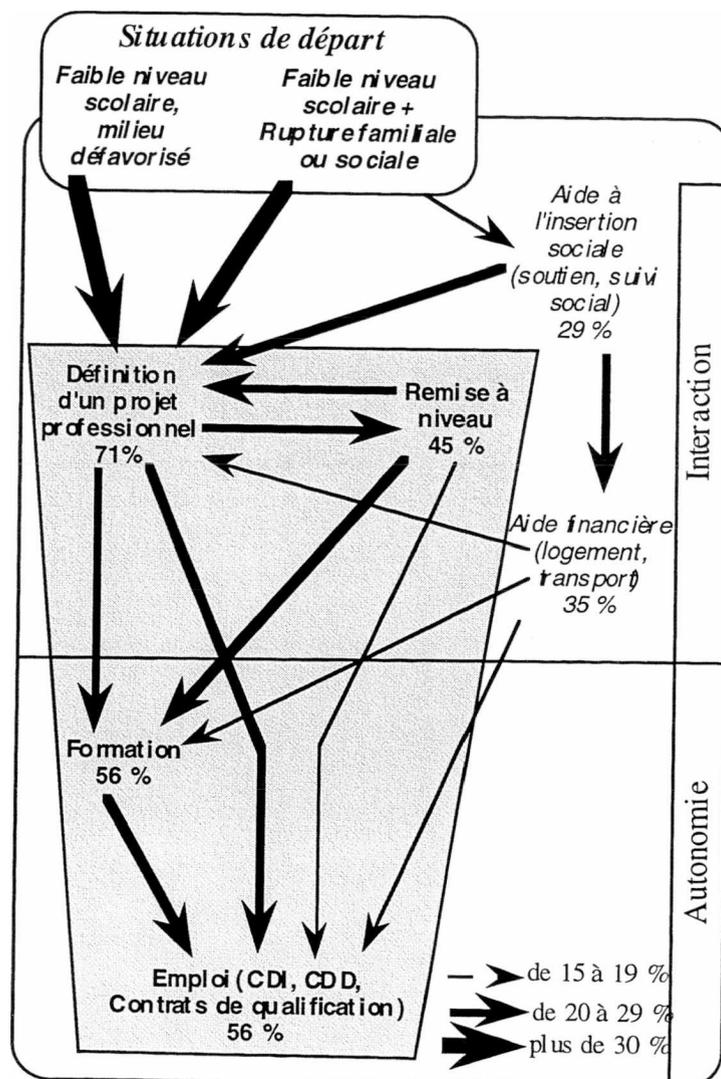
Pour chaque étape, on peut mesurer les flux de transition d'une étape à l'autre. Ainsi, en se situant à une étape déterminée, on peut mesurer le pourcentage de portraits dans lesquels cette étape est suivie par l'étape X, Y ou Z. Le graphe suivant présente les probabilités de passage à partir de l'étape "définition du projet professionnel".



Clef de lecture : 22 % des paragraphes qui décrivent une étape de « définition de projet professionnel » sont suivis par un paragraphe décrivant une étape de « remise à niveau ».

Figure 7 : Graphe de parcours

L'examen de ces différents graphes (Beaudouin et Aucouturier, 1994) montre que pratiquement tous les types de transition sont attestés, qu'il n'y a donc pas, loin s'en faut, un ordre unique dans le déroulement des parcours. Cependant, si l'on choisit de ne tracer que les flux de passage conséquents (les pourcentages supérieurs à 15 %), on voit se dessiner des types de parcours dominants. Ainsi voit-on apparaître une forme de parcours standard : définition du projet professionnel, avec éventuellement une remise à niveau, puis formation et enfin emploi. Dans les portraits, une part importante des jeunes empruntent des détours ou des raccourcis par rapport à ce parcours standard.



Clef de lecture : 71 % des portraits décrivent des jeunes qui sont passés par l'étape de définition d'un projet professionnel. Entre 20 % et 30 % d'entre eux sont allés ensuite directement en remise à niveau, 20 % à 30 % directement en formation et 20 % à 30 % directement en emploi.

Seuls sont représentés les pourcentages supérieurs à 15 %.

Figure 8 : Une très grande variété de parcours

2.3.3. Typologie des parcours

En se fondant sur les couples d'étapes ou variables de transition, on a construit une typologie des parcours. Cette typologie a été obtenue en combinant classification descendante (ALCESTE) et ascendante (SAS) hiérarchique. Nous avons été amené à articuler les résultats de ces deux types de classification, car aucune n'était en elle-même satisfaisante. En effet, la classification descendante avait un gros inconvénient : elle ne permettait pas le classement de parcours très courts, composés de deux ou trois étapes. La classification ascendante a permis l'identification de deux types de parcours très rapides, mais elle regroupait dans la première classe (ci-dessous dénommée "Parcours standard") des parcours "à l'envers" du type "emploi-définition d'un projet"... Sept parcours n'ont été classés ni par la CDH, ni par la CAH.

Le croisement des deux typologies montre qu'en dépit des différences ci-dessus énoncées, on obtient des structures assez semblables par des voies inverses.

CAH \ CDH	Non classés	Parcours standard	Parcours avec remise à niveau	Parcours complexes
Parcours standard	74	117	23	10
Parcours avec remise à niveau	23	3	110	19
Accès direct à la formation	29	0	3	1
Accès direct à l'emploi	28	0	0	2
Parcours complexes	32	1	4	94

Statistic	DF	Value	Prob
Chi-Square	12	636.523	0.000
Likelihood Ratio Chi-Square	12	587.892	0.000
Mantel-Haenszel Chi-Square	1	62.639	0.000
Phi Coefficient		1.054	
Contingency Coefficient		0.725	
Cramer's V		0.609	

Figure 9 : Croisement des deux typologies

Nous avons conservé les trois classes de la CDH (Parcours standard, avec remise à niveau et complexes) et les deux classes de la CAH de parcours rapides (Accès direct à la formation et à l'emploi). Ainsi avons-nous pu réduire un peu le nombre de portraits non classés par la CDH, tout en conservant les classes de la CDH qui nous paraissaient plus cohérentes que celles de la CAH. En fin de compte plus de 20 % des parcours n'ont pas été classés car ils étaient trop atypiques.

- Dans 5 % des portraits, les jeunes accèdent directement à l'emploi sans passer par les étapes de définition du projet ou de formation. Ce sont des parcours réussis et très rapides pour lesquels la Mission Locale ou PAIO joue essentiellement un rôle de "coup de pouce". Ces jeunes n'ont pratiquement jamais été dans des situations initiales de rupture familiale ou sociale.
- Dans 6 % des portraits, les jeunes entrent directement en formation sans passer par l'étape de définition du projet professionnel. Pour la moitié d'entre eux, la formation débouche sur un emploi. Aucun de ces portraits ne décrit des situations sociales et familiales conflictuelles. Il s'agit là encore de parcours simples et rapides. Leur situation finale est tout à fait positive : ils sont plus nombreux que la moyenne à être dans des situations d'emploi et non pas d'emploi en alternance.
- Dans 21 % des portraits, les jeunes suivent un itinéraire proche de la trajectoire standard : *définition d'un projet—formation—emploi*. Les femmes sont plus nombreuses que les hommes à suivre ce type d'itinéraire (17% des hommes , 25% des femmes)
- Pour 25 % des jeunes, cette trajectoire standard doit être complétée par une étape de remise à niveau. C'est après cette étape que les jeunes passent en formation ou trouvent un emploi.
- Enfin, 22 % des portraits correspondent aux parcours les plus complexes où interviennent des problèmes financiers et d'insertion sociale. La majorité de ces jeunes avait de lourds handicaps au départ (conflits familiaux, difficultés d'insertion sociale...). Ces cheminements complexes sont plus fréquents chez les hommes (29 %) que chez les femmes (15 %).

Le graphique suivant représente les résultats de cette typologie.

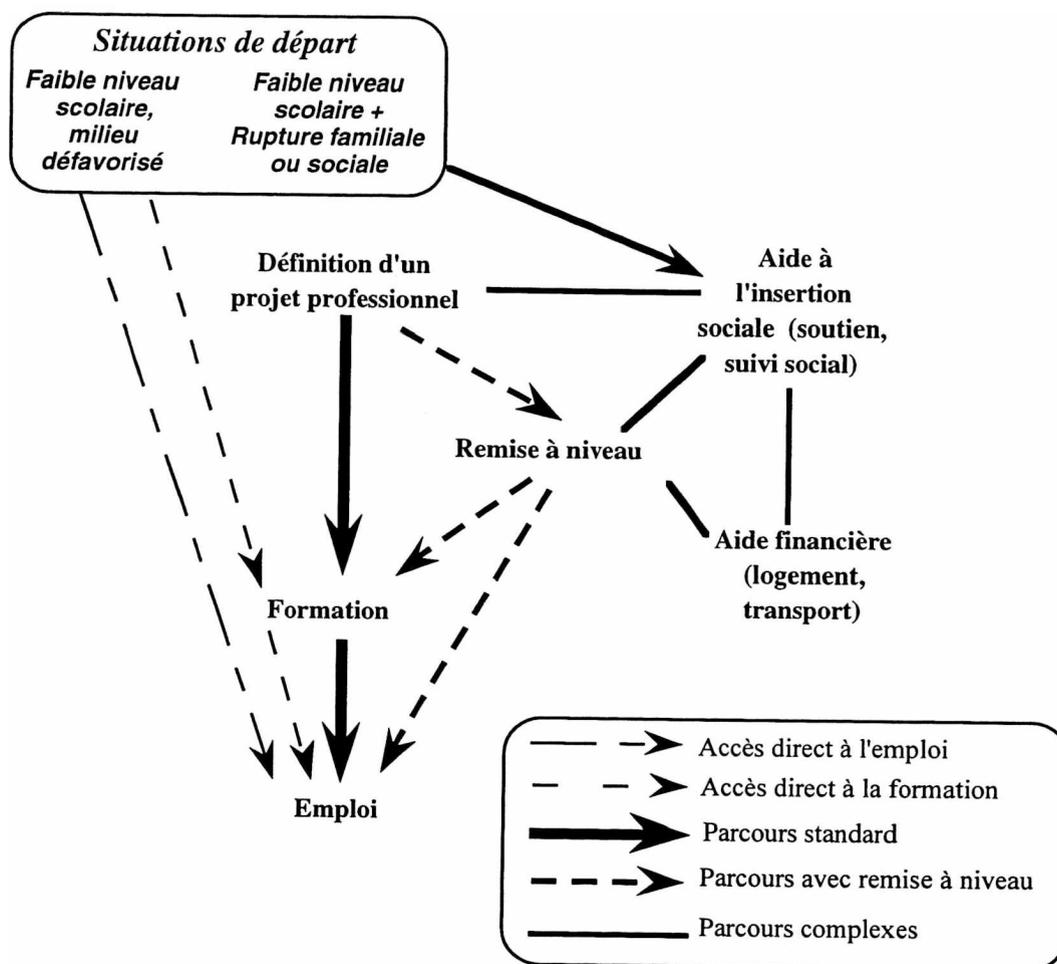


Figure 10 : Typologie des parcours

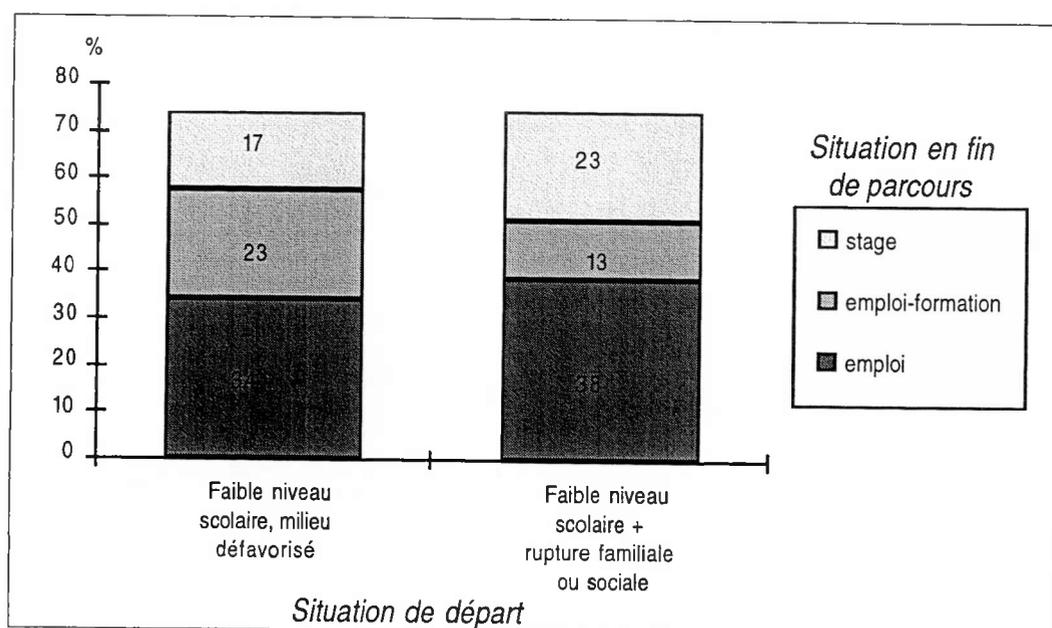
2.3.4. Quelques résultats complémentaires

Comme nous disposons de certaines variables codées « à la main » à partir des portraits il est possible de les mettre en relation avec les types de parcours. Comme il s'agit d'un échantillon non représentatif, aucune conclusion générale ne pourra être tirée des résultats ci-dessous. Ils ne valent que dans le cadre étroit des « jeunes mis en portraits ».

La situation du jeune (par rapport à la formation et à l'emploi) à la fin du portrait a été post-codée. On a ainsi pu comparer la situation finale des jeunes décrits dans les portraits par rapport à leur situation initiale. En fin de parcours, la position professionnelle des jeunes qui au départ étaient en situation de rupture familiale ou sociale s'est sensiblement rapprochée de celle des jeunes qui avaient simplement un faible niveau scolaire. Certes, le profil de fin de

parcours des premiers est un peu différent : ils sont un peu moins nombreux à être en formation en alternance, ce qui aurait pourtant représenté une bonne avancée vers l'insertion professionnelle, mais plus nombreux en stage, ce qui est un état plus précaire. Cependant, la situation sociale et professionnelle de ces jeunes s'est sensiblement améliorée au cours de leur trajectoire par rapport à leur situation initiale.

Des parcours plus complexes —et sans doute plus longs— que les autres leur permettent de surmonter en partie leurs difficultés initiales. Les portraits mettent en évidence une réelle compensation des handicaps de départ.



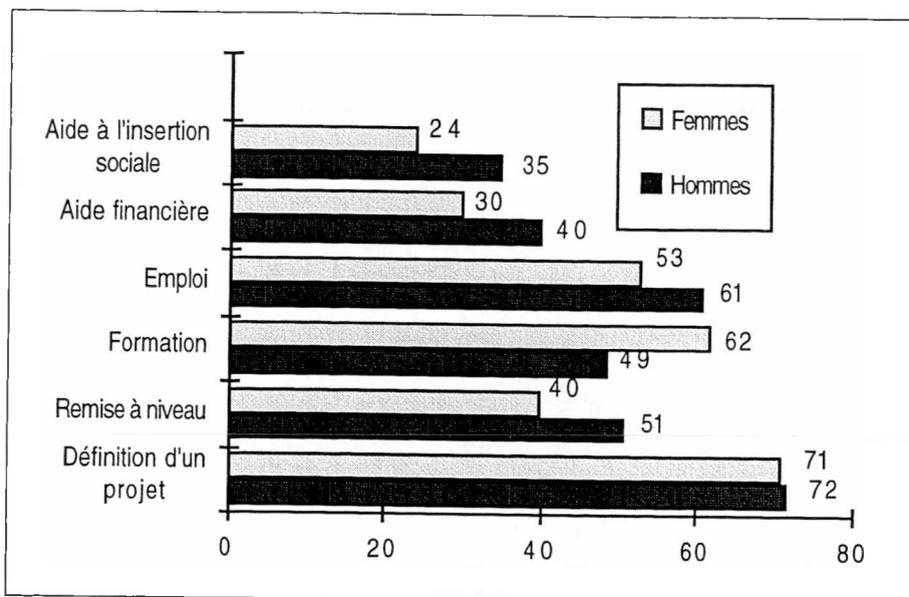
Clef de lecture : 23 % des jeunes qui ont un handicap social au départ (faible niveau scolaire, milieu défavorisé) se retrouvent en formation en alternance à la fin de leur parcours contre 13 % de ceux qui sont au départ en situation de rupture familiale ou sociale.

Figure 11 : Handicap ou traumatisme au départ, des situations similaires en fin de parcours

L'analyse des différences entre hommes et femmes concorde avec des résultats plus classiques. Les femmes sont plutôt orientées vers la formation, les hommes vers l'emploi.

Quand les jeunes sont dans des situations d'urgence, les conseillers ont tendance à les orienter au plus vite vers des emplois qui peuvent se révéler précaires. Or les hommes sont relativement plus nombreux que les femmes à se retrouver dans ce type de situation : ils ont plus souvent besoin au cours de leur parcours de soutien financier ou d'assistance sociale.

Ceci explique sans doute que les hommes soient plus nombreux à passer, souvent à plusieurs reprises, par des périodes d'emploi.



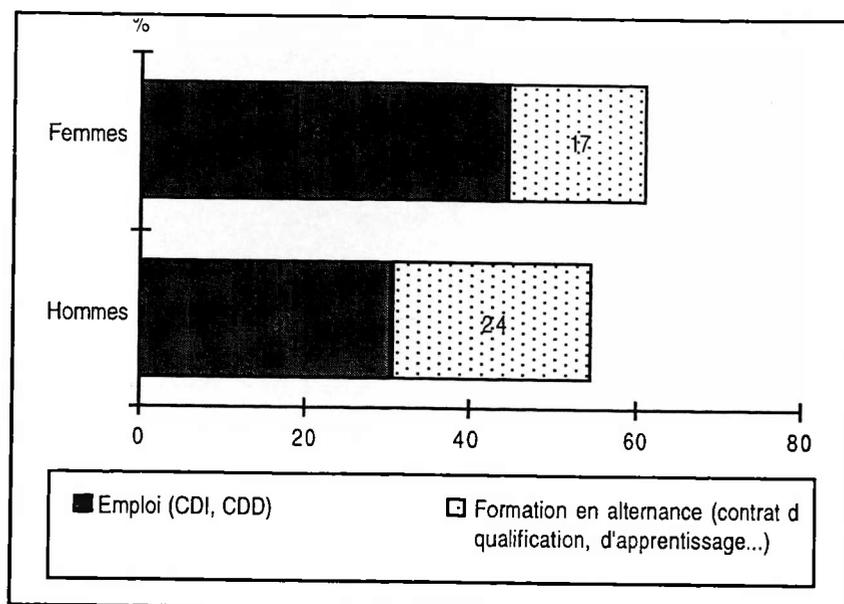
Clef de lecture : dans les portraits, 62 % des femmes sont passées par une étape de formation contre seulement 49 % des hommes

Figure 12 : En cours de parcours, les femmes sont plus nombreuses à passer par la formation, les hommes par l'emploi

Parce que les femmes sont moins souvent dans l'obligation urgente de trouver un emploi — elles ont moins souvent des difficultés financières ou d'insertion sociale—, elles sont orientées vers des modules de formation.

Finalement, la stratégie adoptée pour elles semble plus rentable. En effet, en fin de parcours, les femmes sont nettement plus nombreuses que les hommes à avoir un contrat de travail "classique" (45 % contre 31 %). On comprend mieux pourquoi les conseillers valorisent fortement la formation dans les parcours d'insertion.

Les jeunes hommes, au contraire, orientés le plus vite possible vers l'emploi, sont conduits à des situations finalement plus précaires (stages...).



Clef de lecture : en fin de portrait, 45 % des femmes ont un emploi (CDI ou CDD), contre 31 % des hommes.

Figure 13 : Hommes et femmes en fin de portrait

3. CONCLUSION

Cette étude sur les portraits a permis de trouver une méthode bien adaptée à un corpus particulier : très homogène dans sa forme comme dans son contenu. L'analyse conduit à une représentation synthétique des portraits qui conserve des traces importantes de leur richesse initiale. Elle montre aussi que divers types d'outils permettent mieux qu'un seul de donner une vision fine et appropriée du matériau en question. Ainsi, on a pu voir la complémentarité entre une analyse de flux de passage et une typologie de parcours. La variété des points de vue rend à son objet toute sa complexité tout en permettant une appréhension globale.

Plus généralement, tout discours se présente comme un parcours de lieux lexico-sémantiques qui renvoient à des référents ou à des "visions du monde", pour reprendre l'expression de M. Reinert. Il est donc possible de reconstituer des trajectoires discursives, une fois identifiés ces différents lieux. Reste ensuite à construire l'adéquation entre trajectoire discursive et trajectoire réelle. Parce que le récit traditionnel suit en général la linéarité temporelle, cette opération ne devrait pas toujours poser de trop grandes difficultés. Il nous reste aussi à tester cette méthode sur d'autres types de récits : ceux que l'on obtient par entretiens semi-directifs ou par des "récits biographiques" (Bauer et Maresca, 1992). Ils présentent une difficulté majeure : l'ordre du récit n'est pas toujours cohérent avec le déroulement temporel. Pourtant, l'ordre du discours est toujours signifiant de quelque chose qu'il faudra alors identifier. Il en est de même pour le temps passé à raconter un épisode. La distorsion entre le temps réel et le temps du récit devra être prise en compte comme étant l'expression d'une volonté consciente ou d'un désir inconscient de la part du locuteur.

4. BIBLIOGRAPHIE

- Aucouturier, A. L. (1994). Panels et évaluations des politiques de l'emploi. *Cahier Travail et Emploi*. Paris : La Documentation Française.
- Bauer, D., Maresca, B. (1992). *Lignes de vie : méthodologie de recueil et de traitement des données biographiques*, CRÉDOC, Cahier de recherche n°37, Paris.
- Beaudouin, V., Aucouturier A. L. (1994). *Parcours d'insertion de jeunes en difficulté*. CRÉDOC, Cahier de recherche n°66. Paris.
- Beaudouin, V., Lahlou, S. (1993). *L'analyse lexicale : outil d'exploration des représentations*. CRÉDOC, Cahier de recherche n°48. Paris.
- Benzécri, J.P. (1968). "La place de l'a priori, Organum", *Encyclopedia Universalis*, pp. 11-24.
- Benzécri, J.P., coll. (1981). *Pratique de l'analyse des données, Linguistique et lexicologie*. Paris : Dunod.
- Brunet, E. (1992). *Hyperbase*, CUMFID n°17 (version 1.5), 84 p.
- Brunet, E. (1993). "Un hypertexte statistique : Hyperbase". In S. J. Anastex (eds.), (1993). *JADT 1993*. Paris : TELECOM.
- Constant, P. (1991). *Analyse syntaxique par couche*. Paris : Doctorat ENST.
- Diday, E. et Lebart, L. (1977). "L'analyse des données". *La recherche*, n°74, pp. 15-25.
- Genette, G. (1972). *Discours du récit. Figures III*. Paris : Seuil, Collection Poétique, pp. 67-282.

- Hamburger, K. (1986 pour la traduction française). *Logique des genres littéraires*. traduction P. Cadiot, Préface G. Genette. Paris : Seuil, Collection Poétique, 1957.
- Lahlou, S. (1995). *Penser Manger : les représentations sociales de l'alimentation*, Thèse EHESS sous la direction de Serge Moscovici.
- Lebart, L., Salem, A. (1988). *Analyse statistique des données textuelles*. Paris : Dunod.
- Lebart, L., Salem, A. (1994). *Statistique textuelle*, Paris : Dunod.
- Maranda, P. (1978). "The Popular Subdiscourse: Probabilistic Semantic Networks". *Current Anthropology*, 19, pp.396-397.
- Maranda, P. (1989). *Discan. User's Manual*, Québec, Canada.
- Maranda, P. (1992). "Mother Culture Is Watching Us : Probabilistic Structuralism". In : Nardocchio, Elaine (ed.), *Reader Response to Literature. The Empirical Dimension*. Berlin : Mouton de Gruyter, p. 173-192.
- Maranda, P. et Nze-nguema, F. -P. (1994). *L'unité dans la diversité culturelle. Un geste bantu*. Québec : Presses de l'Université Laval, 232 p.
- Markov, A. A. (1913). Un exemple de recherche statistique sur le texte d'*Eugène Onéguine* illustrant la liaison des épreuves en chaîne. *Bulletin de l'Académie impériale des sciences*, 6è série, t. 7, pp. 153-162.
- Martin, O. (1995). *La mesure en psychologie : au-delà de l'outillage formel. Recherches historiques sur les contributions des outils statistiques à la construction des faits psychologiques*. Thèse EHESS sous la direction de B. Matalon.
- Ourtau, M., Werquin, P. (eds.), (1994). *L'analyse longitudinale du marché du travail*. Marseille : CEREP, n°99.
- Petruszewycz, M. (1983). "Chaînes de Markov et statistiques linguistiques". *Mots*. Paris : Presses de la fondation nationale des sciences politiques, n° 7, pp. 85-95.
- Pigamo, F. (1990). *Outils de traitement sémantique du langage naturel*. Paris : Doctorat ENST.
- Reinert, M. (1983). "Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte". *Les cahiers de l'analyse des données*, Vol VIII, n° 2.

- Reinert, M. (1990). ALCESTE : "Une méthodologie d'analyse des données textuelles et une application : Aurélia de Gérard de Nerval". *Bulletin de méthodologie sociologique*, n°26, pp. 24-54.
- Reinert, M. (1993). "Les "mondes lexicaux" et leur logique". *Langage et société*, Maison des Sciences de l'Homme, n°66, p. 5-39.
- S. J. Anastex (eds.), (1993). *JADT 1993 : Actes des Seondes journées internationales d'analyse statistique de données textuelles*. Paris : TELECOM.

5. TABLE DES ILLUSTRATIONS

<i>Figure 1 : Analyse de récit</i>	12
<i>Figure 2 : Exemple de portrait</i>	16
<i>Figure 3 : Identification des étapes possibles</i>	18
<i>Figure 4 : Flux de passage et types de parcours</i>	20
<i>Figure 5 : Regroupements effectués à la main</i>	24
<i>Figure 6 : Situations de départ</i>	26
<i>Figure 7 : Étapes de parcours</i>	27
<i>Figure 8 : Graphe de parcours</i>	30
<i>Figure 9 : Une très grande variété de parcours</i>	31
<i>Figure 10 : Croisement des deux typologies</i>	32
<i>Figure 11 : Typologie des parcours</i>	34
<i>Figure 12 : Handicap ou traumatisme au départ, des situations similaires en fin de parcours</i>	35
<i>Figure 13 : En cours de parcours, les femmes sont plus nombreuses à passer par la formation, les hommes par l'emploi</i>	36
<i>Figure 14 : Hommes et femmes en fin de portrait</i>	37

Dépôt légal : Novembre 1995

ISSN : 1257-9807

ISBN : 2-84104-049-6

CAHIER DE RECHERCHE

Récemment parus :

Le petit déjeuner : Représentations et comportements

Valérie BEAUDOUIN, Aude COLLERIE DE BORELY, Marie-Odile GILLES, Saadi LAHLOU, Franck MARSAL, Laurence PEYRE, Stéphanie WATTRELOT et François YVON - n°75 (1995)

Pauvretés, ruptures et évaluations de politiques sociales

Département Évaluations des Politiques Sociales du Crédoc - n°76 (1995)

L'évaluation des politiques publiques et le concept de "bénéficiaires"

Guy POQUET - n°77 (1995)

L'aire et la manière

Patrick DUBÉCHOT, Marie-Odile GILLES et Michel LEGROS - n°78 (1995)

Restauration hors foyer et nutrition

Gloria CALAMASSI-TRAN, Pascale HÉBEL et Patrice LE FRANCOIS - n°79 (1995)

Évaluer ou les efforts de la Métis

Michel LEGROS - n°80 (1995)

L'analyse de réseau comme approche des processus d'interaction des acteurs sociaux

Bruno MARESCA et Elisabeth SIÉCA - n°81 (1995)

Président : Bernard SCHAEFER Directeur : Robert ROCHEFORT
142, rue du Chevaleret, 75013 PARIS - Tél. : (1) 40.77.85.00

ISBN : 2-84104-049-6

CRÉDOC

Centre de recherche pour l'Étude et l'Observation des Conditions de Vie