DUALITÉ ENTRE ESPACE DES VARIABLES ET ESPACE DES SUJETS EN ANALYSE STATISTIQUE IMPLICATIVE

Régis GRAS[1], Dominique LAHANIER-REUTER[2]

TITLE

Duality between variables space and subjects space of the statistic implicative analysis

RÉSUMÉ

L'Analyse Statistique Implicative vise à quantifier, à travers les concepts de typicalité et de contribution, la qualité de la relation de chaque sujet à l'élection de chemins du graphe implicatif ou à celle de classes des hiérarchies de similarité et de cohésion implicative. En retour, ces expressions quantitatives induisent sur l'ensemble des sujets une structure métrique qui permet de comparer les positions respectives des sujets par rapport aux concepts extraits de la base de données initiale. En particulier, il est possible de faire apparaître les sujets « atypiques », c'est-à-dire extrêmes, à l'égard de ces concepts. Une application portant sur des élèves de classes élémentaires examine les façons qu'ils ont d'identifier les différentes disciplines qui leur sont enseignées.

Mots-clés *: Contribution, graphe implicatif, hiérarchie cohésitive, intensité d'implication, règle, sujet atypique, typicalité.*

ABSTRACT

The Implicative Statistics Analysis tends to quantify, through the concepts of typicality and contribution, the quality of the relationship that links each subject to the implicative graph ways or to the classes for similarity hierarchies and to the implicative cohesion ones. These quantitative formulas conversely provide a metric structure for the subjects set, allowing comparison of the places of subjects according to the concepts extracted from the first data base. Eventually borderlines subjects may be detected, that is the furthest ones. We use this approach in exploring how elementary school students identify the different subjects they are taught

Key-words*. Contribution, implicative graph, cohesitive hierarchy, rule, borderline subject, typicality.*

1 Introduction et problématique

1.1 Rappels

L'Analyse Statistique Implicative (A.S.I.) est une méthode d'analyse de données qui opère sur un tableau de valeurs numériques non négatives (Gras et al, 2001a). Elle

[1] Ecole Polytechnique de l'Université de Nantes, Equipe Connaissance et Décision, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241, E-mail : regisgra@club-internet.fr, http://math.unipa.it/~grim/homegras_03.htm

[2] Equipe Théodile CIREL, Université Lille 3, Villeneuve d'Ascq, France, E-mail: dominique.lahanier@univ-lille3.fr

croise un ensemble E de sujets en nombre n et un ensemble V de variables binaires (attributs, par ex.), numériques (modales, par ex.), intervalles (floues, par ex.) ou vectorielles (suites, par ex.)). Ces variables sont dites **principales ou actives** car elles-seules interviennent directement dans les calculs de leurs relations mutuelles. Elles reflètent le comportement ou l'attitude des sujets alors que des descripteurs (**objectifs**) de ces sujets sont dits **supplémentaires** et informent sur leur identité ou leur appartenance à des groupes déterminés.

L'A.S.I. a pour objectifs :

- **extraction de règles** entre les variables principales de la forme « a implique b » (on note a=>b) dès lors que l'observation de la variable a s'accompagne généralement de l'extraction de règles entre les variables principales de la forme « a implique b » (on note a=>b) dès lors que l'observation de la variable a s'accompagne généralement de celle de la variable b ;
- **quantification de la qualité de la règle a⇒b** en rapport avec l'étonnement statistique de constater un certain nombre (faible par ex.) de contre- exemples à cette règle alors qu'en toute hypothèse les variables seraient indépendantes ; autrement dit, la règle n'est pas nécessairement stricte mais partielle (quasi), ce qui est fréquemment le cas en sciences humaines ; deux valeurs modélisent cet étonnement statistique : **l'intensité d'implication** dite **classique** et **l'intensité d'implication entropique,** notées ici l'une ou l'autre $\psi(a,b)$ qui est d'autant plus proche de 1 que l'étonnement statistique est important ou que l'entropie est faible ; c'est un indice de **qualité** de la règle en un certain sens prédictif ;
- **structuration de l'ensemble des règles** dont les formes graphiques s'expose selon un graphe (**graphe implicatif**) et un arbre hiérarchique orienté (**hiérarchie cohésitive**).

Sur la base de la connaissance de la satisfaction graduée de *chaque sujet x de E* à chacune des règles i, l'A.S.I. quantifie la qualité de cette relation selon deux paramètres (Gras et Régnier, 2009) :

- d'une part **la distance de typicalité** de x par rapport à un chemin C du graphe ou une classe C de la hiérarchie, distance notée d(x,**C**), prend en compte la qualité relative $\psi_{x,i}$, de la relation du sujet x à la règle i. La typicalité est définie à partir d'une distance de type χ^2 [3] entre les deux vecteurs $(\psi_{x,1},\ldots, \psi_{x,g})$ et $(\psi_1,\ldots,\psi_g)$:

$$d(x,\mathbf{C})=\left[\frac{1}{g}\sum_{i=1}^{i=g}\frac{[\psi_i-\psi_{x,i}]^2}{1-\psi_i}\right]^{\frac{1}{2}} \quad \mathbf{(1)}$$

*où ψ_i est l'intensité d'implication maximale des règles au sein des sous-chemins i du chemin C du graphe ou des sous-classes i de la classe C de la hiérarchie et

[3] Il en résulte que la différence relative entre la qualité de la relation de x à la règle i et la qualité de cette règle est d'autant plus grande que cette dernière l'est également. C'est un moyen de renforcer cette « distorsion ».

* où g est le nombre de sous-chemins de **C** ou le nombre de sous-classes de **C**. Ce nombre, qui vérifie formellement les trois axiomes d'une distance, n'est autre également que la distance du type χ^2 entre les deux distributions $\{1-\varphi_i\}$ et $\{1-\varphi_{x,i}\}$ pour i=1 à i=g, qui expriment les écarts entre les implications génériques de C contingentes et l'implication stricte. La distance d_C permet de conférer à E une d_C-structure topologique discrète. Le sujet x_0, éventuellement fictif, à distance nulle de C est appelé sujet optimal ;

- d'autre part **la distance de contribution** qui prend en compte cette qualité si la règle était stricte. Elle est définie à partir de la distance euclidienne :

$$d(x,\mathbf{C})=\left[\frac{1}{g}\sum_{i=1}^{i=g}\left[1-\psi_{x,i}\right]^2\right]^{\frac{1}{2}} \quad \mathbf{(2)}$$

Les propriétés de cette distance sont les mêmes que celles de la distance de typicalité.

Enfin, nous définissons la mesure de typicalité à partir du rapport entre la distance de typicalité relative au sujet considéré et la distance à C, classe ou chemin, la plus grande dans l'ensemble des sujets. Cette distance maximale est celle des sujets y dont les $\psi_{y,i}$ sont tous nuls ou très faibles. Ces sujets sont donc ceux les plus opposés aux règles génériques. La typicalité d'un sujet est alors d'autant plus grande qu'il s'écarte de ces mêmes sujets, donc qu'il manifeste un comportement comparable à celui du sujet théorique optimal. La typicalité d'une catégorie de sujets ou d'une variable supplémentaire G[4] s'en déduit alors :

La typicalité du sujet x à la classe ou au chemin C est mesurée par :

$$\gamma(x,C)=1-\frac{d(x,C)}{\max_{y\in E}\{d(y,C)\}}$$

De la même façon, l'A.S.I. définit la typicalité et la contribution des variables supplémentaires à un chemin du graphe ou à une classe de la hiérarchie.[5]. On dira par exemple, que les élèves d'une classe de type « Freinet » sont les plus typiques de la règle où s'enchaînent les matières dont l'importance subordonne la réussite en français : Grammaire => Orthographe => Conjugaison.

La typicalité de la variable supplémentaire G à la classe ou au chemin C est mesurée par :

$$\gamma(G,C)=\frac{1}{cardG}\sum_{x\in G}\gamma(x,G)$$

Remarque 1

Les distances (**1**) et (**2**) constituent aussi des **profils** du sujet x selon la variable **C**. Bien que non normalisées, ces profils peuvent l'être si nous envisageons, par exemple, une étude factorielle, de type A.F.C. ou A.C.P., portant sur le croisement de vecteurs

[4] Les deux mots « catégorie » et « variable supplémentaire » seront utilisés indifféremment, le premier ayant une charge sémantique plus forte que le second.

[5] Le logiciel CHIC (Couturier R., 2008) permet de traiter les données afin d'obtenir les représentations des structures engendrées par la méthode implicative. En outre, il permet de représenter la hiérarchie des similarités définie initialement par I.C.Lerman et revue par Maria Polo dans sa thèse (Polo M., 1996) selon un critère de cohésion afin d'éviter la formation de classes incohérentes.

profils de l'ensemble E avec l'ensemble des chemins ou des classes de variables tels que **C**. Nous ne prolongerons pas la recherche présente dans cette direction qui restera ouverte, mais il serait loisible de l'envisager lors d'une prochaine recherche, d'autant que le point suivant nous y engagerait.

1.2 Problématique

Les typicalité et contribution d'un sujet quelconque de E, espace des sujets, nous permettent de définir sur E, relativement au chemin ou à la classe **C**, une topologie discrète d'espace normé. Une norme est associée à la distance de type χ^2 entre deux sujets x et y donnée par la relation suivante (voir la note précédente) où ψ x,i et ψ y,i sont respectivement les typicalités (ou les contributions) des sujets x et y à chacune des g règles i contenues dans **C**

$$d_C\,(x,y) = \left[\frac{1}{g}\sum_{i=1}^{i=g}\frac{\left[\psi_{x,i}-\psi_{y,i}\right]^2}{1-\psi_i}\right]^{\frac{1}{2}} \quad \textbf{(3)}$$

Notons bien que, le chemin C ou la classe C de variables étant donnés, cette distance est une fonction de C. Elle établit une première correspondance entre l'ensemble des sujets et l'ensemble des variables structuré par un graphe implicatif ou une hiérarchie cohésitive, restreints à C. On voit alors que la distance de typicalité donnée plus haut n'est que la spécification de d_C aux sujets respectivement x et x_0 (sujet optimal). La distance d_C permet de conférer à E une d_C-structure topologique discrète.

Une application intéressante peut consister à déterminer le ou les sujets appartenant à une boule de diamètre donné et de centre l'un des sujets pré-désignés.

Par une procédure de type « agrégation autour des centres mobiles », par exemple la méthode des « **nuées dynamiques** » de E.Diday (Diday E., 1972), on peut alors faire une partition $\mathcal{P}$ sur E relative au concept expliquant la classe ou le chemin **C**. Cette partition $\mathcal{P}$ peut être comparée à celle que définiraient la ou les variables descripteurs auxquelles x et y satisfont. On pourra utiliser, pour ce faire, le logiciel CHIC, avec **l'option intervalles**, où deux variables ou plus figureront dans le tableau : d'une part les valeurs de la contribution ou la typicalité de chaque sujet à l'élection de la classe ou du chemin **C** et, d'autre part, les valeurs d'appartenance aux variables descripteurs des sujets. Il sera ainsi possible d'étudier la relation (de similarité ou implicative) de la partition $\mathcal{P}$ et de la ou des autres partitions définies avec les descripteurs.

A travers cet objectif, nous observons la propriété épistémique de l'**émergence** d'un « *tout* » en tant que *système*, plus riche que la somme de ses « *parties* ». C'est le cas d'une nouvelle structure - celle des sujets- qui naît, qui *émerge* de la prise en compte de phénomènes –des règles entre variables- engendrées par ces mêmes sujets. Son sens et son efficience sont comme mis en scène par cette nouvelle architecture comme le sont déjà les autres structures révélées sur l'ensemble des variables : le graphe implicatif et la hiérarchie cohésitive.[6]..

[6] Citons, à ce sujet, deux extraits du chapitre de L.Sève (Sève et al, 2005, p.58-59) : « …le tout ne se compose de rien d'autre que de ses parties, et pourtant il présente en tant que tout des propriétés n'appartenant à aucune de ses parties. Autrement dit, dans le passage non additif, non linéaire des parties

Remarque 2

C'est pour cette raison que nous évoquons l'extension de la notion de **dualité** entre les deux espaces V et E en jeu, sujets (et descripteurs en variables supplémentaires) et variables, comparable à celle dont nous parlerions dans le domaine des espaces vectoriels où s'échangent des structures**.** En l'occurrence dans l'A.S.I.**,** la structure de treillis du graphe implicatif et la structure de hiérarchie orientée[7] de l'arbre cohésitif confèrent à l'ensemble des sujets (et de leurs descripteurs) une double structure métrique :

- d'une part entre les sujets eux-mêmes, leur distance étant un reflet de leur comportement plus ou moins voisin aux chemins du graphe ou aux classes de hiérarchie ;
- d'autre part entre les sujets et les éléments structurés du graphe (chemins du treillis) et de l'arbre (classes de la hiérarchie orientée).

On établit ainsi une véritable symbiose entre les deux espaces : les modulations sur l'un (par exemple, un changement de seuil sur le graphe) se traduisent par des frémissements sur l'autre. Leurs structures formalisées permettent de considérer l'ensemble des sujets de façon dynamique et non plus comme résultant d'un transport de structure, c'est un transfert de dynamique de type inclusif ou emboîtement. Ces propriétés doivent être exploitées par l'analyste s'il veut donner un sens enrichi aux relations établies sur les deux espaces. C'est d'ailleurs l'objectif premier de l'application qui suivra.

Remarque 3

A l'instar de la **remarque 1**, nous observons que la distance (**3**) dénote la différence des profils des deux sujets x et y selon la variable **C**. Nous retrouvons ainsi les concepts et les conditions, à la normalisation près, qui permettent les traitements factoriels classiques, associant ainsi opportunément notre approche classificatoire A.S.I. et celle de l'A.F.C. ou de l'A.C.P. Nous ne renonçons pas à comparer ces deux approches comme il a été dit plus haut.

Remarque 4

Les propriétés relatives à la dualité des espaces V et E que nous venons de rappeler ne sont pas les seules qui spécifient l'A.S.I. par rapport à d'autres mesures de règles d'association (Gras et Couturier, 2011) et nous les exploiterons. Citons par exemple, la non symétrie de l'indice d'implication, la prise en compte statistique des contre-exemples à la règle et d'une grande variété de variables, l'extraction des cas rares, la décroissance de la mesure avec la trivialité, la conjugaison avec la contraposée de la règle, la variété des représentations graphiques, etc..

---

au tout, il y a *apparition de propriétés* qui ne sont d'aucune manière *précontenues* dans les parties et ne peuvent donc s'expliquer par elles »…« Tout se passe donc comme si se produisait une *génération spontanée* de propriétés du tout…C'est le paradoxe de **l'émergence**. »

[7] Nous avons démontré que l'arbre cohésitif orienté par les méta-règles était bien une hiérarchie au sens de la théorie des graphes du fait de la distance ultramétrique dont elle est munie (Gras et Kuntz, 2008)

1.3 Application des notions de typicalité et de contribution[8] en A.S.I. pour extraire les individus «atypiques[9] ».

Une telle application présente deux objectifs **:**

1. Extraire les individus qui relativement à une classe d'une hiérarchie (similarités ou cohésitive) ou un chemin (implicatif) **C** contribuent le plus (resp. le moins) à cette classe ou ce chemin. Il suffira de considérer les valeurs des contributions indiquées par CHIC. Si l'on peut conceptualiser cette classe ou ce chemin, on dira ici que ces individus sont extrêmes ou atypiques, voire selon le vocable actuel, « atypiques » par rapport au concept associé. Ce sont des individus « en excès » ou positifs (resp. « en défaut » ou négatifs) par rapport au concept : ils en sont une image forte ou respectivement une image antithétique. En didactique par exemple, cette extraction permet de dépasser l'examen seul des résultats ou bien globaux d'élèves (performance totale) ou bien spécifiques d'une partie d'épreuve (item d'un test). En revanche, c'est par rapport à l'organisation structurée des variables que sera pratiquée cette détermination de sujets « extrêmes ». Ainsi, on dira, par exemple, que tel élève contribue le plus (resp. le moins) à une règle qui indique la réussite à telle ou telle relation positive entre deux items ;
2. Constituer des ensembles d'individus, atypiques négatifs (resp. positifs) à l'aide d'un algorithme du genre « nuées dynamiques » de (Diday, 1972) (cf.§ 1-2) et comme nous l'avons fait pour la recherche de maximisation des sous-intervalles d'un intervalle donné (Gras et al., 2001b). On pourra utiliser pour cela les valeurs obtenues pour chaque individu (en choisissant librement un nombre de sous-intervalles) ou bien les typicalités ou bien les contributions ou les valeurs déduites des distances entre individus définies par les mesures de type $\chi 2$ des notions de typicalité ou de contribution. Cette dernière méthode permet de construire une véritable topologie métrique sur l'ensemble des individus.

Supposons un chemin implicatif ou une classe cohésitive choisis, porteurs de sens conceptuels et les intervalles créés par la méthode des nuées dynamiques. Les sujets « atypiques » appartiennent respectivement aux premier et dernier sous-intervalles.

En outre, la formule (**3**) : $d_C\ (x,y) = \left[\frac{1}{g}\sum_{i=1}^{i=g}\left[\psi_{x,i} - \psi_{y,i}\right]^2\right]^{\frac{1}{2}}$ permet de définir par exemple une « boule » centrée en x et dans laquelle on retrouve les sujets atypiques les plus proches de x.

Nous présentons en annexe une autre application des propriétés métriques que nous venons de mettre en évidence. Nous nous contentons pour l'instant d'indiquer la

[8] Dorénavant, sauf en cas d'ambiguïté des deux notions, nous désignerons sous le vocable « contribution » aussi bien la notion que nous avons rappelée que celle de typicalité également définie plus haut.

[9] Dans le jargon anglophone, en Intelligence Artificielle et Fouille de données, on emploie fréquemment l'expression « borderlines » pour qualifier des sujets qui conduisent à des situations limites par rapport au comportement de la majorité de la population qui alimente les données

démarche méthodologique envisagée qui conduirait à une nouvelle classification des sujets intégrant leur sensibilité à certaines règles dont la sémantique serait claire.

2 Application à l'analyse d'un questionnaire d'élèves et la recherche d'élèves extrêmes (atypiques)

2.1 Problématique et cadre de la recherche

Le corpus que nous nous proposons d'explorer dans cette optique de recherche a été constitué d'après l'étude des réponses à un questionnaire. Ce dernier a été distribué en trois fois à 303 élèves de CM1(10-11 ans) et CM2[10](11-12 ans) durant les deux années scolaires 2007-2008 et 2008-2009. Il constituait l'un des éléments de la méthodologie associée à un projet de recherche en didactiques disciplinaires. Cette étude a duré quatre ans et rassemblé plusieurs didacticiens de différentes disciplines : Français, Oral, Histoire-Géographie, Sciences et Mathématiques. Ce projet tendait à comprendre et décrire les façons dont des élèves de fin d'école primaire identifient certaines des disciplines qui leur sont enseignées. En effet, des résultats de recherche précédents (Clément 2005, Cohen-Azria, Hassan et Tutiaux-Guillon, 2011, Kus, 2006, Lahanier-Reuter et Reuter, 2011, Lécuyer,2005, Marhem, 2006, Ségismont, 2006, Van Meenen, 2006) ont pu montrer à quel point cette reconnaissance pouvait être malaisée, en tous les cas très variable dans son acuité et sa pertinence selon les élèves et selon les disciplines considérées. Discerner les spécificités des espaces disciplinaires est un travail « caché » en ce qu'il est parfois peu explicitement programmé et évalué. Il n'en demeure pas moins que ce discernement se construit au travers de multiples dimensions, telles celles des particularités des contenus enseignés, de leur hiérarchie et de leur structure, des tâches et des conduites qui y sont attendues, autorisées et interdites, des modes de lecture et d'écriture recommandées et prescrites. Il n'est donc guère surprenant que la réussite et l'échec disciplinaires semblent bien liés à la pertinence et à l'adéquation de ces discernements avec ceux que préconise le monde scolaire.

Pour rendre compte de ces discernements et de ses multiples dimensions, Reuter (2007a) avance le concept de « conscience disciplinaire ». Le degré de conscience disciplinaire, essentiellement variable parmi les élèves et au long de la scolarité, est en conséquence à prendre en compte pour comprendre des réussites et des échecs particuliers dans les différentes disciplines. Nous supposons que ce degré de conscience disciplinaire est fonction de la discipline interrogée, mais nous ne savons guère de choses sur ces relations : sont-elles indépendantes totalement ? Est-il envisageable qu'un élève construise l'espace de chaque discipline de façon absolument indépendante ? Nous supposons également que ce degré de conscience disciplinaire dépende de la façon dont ces espaces et leurs spécificités sont élaborés dans les classes que l'élève fréquente. Pour le dire autrement, les aspects didactiques des enseignements sont, selon nous, liés aussi à ces degrés de conscience. En revanche, nous ignorons encore comment ces aspects peuvent moduler ces discernements. Enfin, l'école, la classe, ne sont sans doute pas les seuls espaces sociaux où les différents univers disciplinaires sont convoqués : les discours des parents, des frères, sœurs et copains,

[10] Il comporte en fait une vingtaine de questions, mais nous ne prendrons en compte ici que 4 d'entre elles. Il s'agit essentiellement de questions ouvertes.

leurs pratiques, leurs conseils etc. participent sans doute aussi de cette élaboration. Encore une fois, ces interactions nous sont mal connues.

Le projet de recherche était donc celui de décrire des degrés différents de conscience disciplinaire, dans quatre disciplines scolaires (Français, Mathématiques, Histoire-Géographie, Sciences) à la fin de l'école primaire, et secondairement, de tenter de comprendre les relations qui les lient – sans doute- aux disciplines qui en sont l'objet ainsi qu'aux modes d'enseignement et aux milieux socio-culturels des élèves. Douze classes de CM1 et CM2 ont alors été choisies : elles se différencient par les milieux socio-culturels de leurs élèves, par les choix et inscriptions pédagogiques des maîtres. Les variables supplémentaires du corpus étudié sont par conséquent les traces de ces choix : elles concernent tout d'abord les douze classes (et en conséquence les douze maîtres qui en sont responsables), les milieux (de très favorisé à très défavorisé) et les inscriptions revendiquées des maîtres dans des modes pédagogiques (de la pédagogie de type Freinet à l'absence d'inscription[11]). Ainsi ces variables supplémentaires seront utilisées plus loin pour interpréter si possible les positions des sujets extrêmes.

2.2 Les variables en jeu

Nous ne retenons dans ce cadre que deux disciplines scolaires : le français et les mathématiques. Nous savons déjà que ces deux disciplines sont le plus souvent reconnues par les élèves, de façon relativement contrastée sur deux dimensions, celle de l'écriture et celle du type de textes à écrire/à lire[12]. Les questions du questionnaire informé par les élèves que nous avons choisi d'explorer ici sont les suivantes :

1. Qu'est-ce que tu as appris d'important cette année en Français ?
2. Qu'est-ce que tu as appris d'important cette année en Mathématiques ?
3. Qu'est-ce qui est important selon toi, pour réussir en Français ?
4. Qu'est-ce qui est important selon toi, pour réussir en Mathématiques ?

A ces questions ouvertes, les élèves ont répondu de façon relativement différente. C'est bien là l'intérêt de notre étude, que d'essayer, au travers de la diversité des discours, de saisir les diversités des formes de conscience disciplinaire et d'en établir, même si le terme est malencontreux, des degrés. Nous sommes, rappelons-le, à la recherche des modes de discernement des disciplines, et à la recherche de leurs différentes organisations.

De façon différente disons-nous, mais cette diversité n'est pas immense. Des fragments de discours, des expressions sont récurrentes : « bien écouter l'enseignant [Madame] [La maîtresse] » aux deux dernières questions, « [bien] [mieux] calculer » à celles qui concernent les mathématiques par exemple. Ce sont ces fragments de textes qui *sont les bases* des variables décrivant le corpus : par exemple, la variable MICo

[11] Ce que nous interprétons comme une inscription dans une pédagogie de type « classique ».

[12] Pour le dire très vite, les élèves conçoivent relativement tôt et massivement qu'il faut « faire court » en mathématiques et « long » en français, puisque dans la première « le fond est plus important que la forme » et inversement pour la seconde ; que certains types de textes – le narratif et le descriptif en particulier- sont récurrents et objets d'étude en français, tandis qu'ils sont pratiquement absents en mathématiques.

désigne la présence/absence dans la réponse à la question « Qu'est-ce qui est Important pour toi pour réussir en Mathématiques » de l'énoncé « [savoir] compter ».

Cependant, ce découpage et ce relevé de fragments de textes ne sont empiriques que pour partie ; ils ne sont pas entièrement soumis aux données recueillies. Ils sont également soutenus par des décisions qui relèvent du cadrage théorique de notre projet. Ainsi, nous avons considéré comme fragment à relever *poser des divisions* dans la réponse de Jordan, CM2, Chopin (nom de classe) à la deuxième question « On apprend à compter et à calculer, à poser des divisions. », mais *divisions, soustractions, additions,* dans la réponse de Thomas « Les divisions, les soustractions et les additions », dans la même classe, à la même question. Le premier fait référence à une technique opératoire et à sa maîtrise, tandis que le second liste des termes. L'identification des contenus « importants » à apprendre n'est pas identique : d'un côté un contenu nouveau en CM2, réputé difficile, explicitement désigné comme un enjeu de savoir et d'évaluation, de l'autre des « objets » énumérés, qui peuvent aussi bien désigner des objets discursifs (on en parle), visuels (il y en a au tableau, sur les fiches) ou objets d'étude. L'identification des « contenus appris », la hiérarchisation de ces derniers est différente. En conséquence, elles ne procèdent pas de la même façon d'envisager l'ensemble des contenus disciplinaires, dans le contexte de la réponse à ce questionnaire.

2.3 Traitement et analyse des données

Les variables relevées sont en conséquence nombreuses, nous ne retenons ici que celles qui interviennent dans des chemins révélés par l'analyse implicative. Elles ont été identifiées à partir des réponses aux questions ouvertes et traduites sous forme binaire à travers différentes modalités de réponse. Le tableau ci-dessous récapitule ces variables, leurs éléments de codage et leur signification.

TABLEAU 1 – *Légende et codage*

MR : Maths réussite	FR : Français réussite	MI : Maths appris Important	FI : Français appris Important
Ca : Calculs	At : Attitude	Ac : Activités spécifiques	A Autres
Co : Compter	C : Conjugaisons	Au : Autres	C Conjugaisons
E Écouter	Ec : Écouter	Ca : Calculs	Ecr : Écrire
L : Leçons	Ecr : Écrire	Co : Compter	G : Grammaire
O : Opérations	G : Grammaire	Di : Division	LI : Lire
Rs : Règles scolaires	L : Leçons	O : Opérations	O : Orthographe
	LI : Lire		T : Terminaisons
	O : Orthographe		V : Verbes
	Rs : Règles scolaires		
	V : Verbes		

Nous choisissons comme méthode d'analyse de données l'analyse statistique implicative (A.S.I.) (Gras, 96, 2001a, 2008)en vue de mettre en évidence une structure non symétrique des relations entre les variables et d'éventuelles causalités ou subordination entre elles. Le traitement automatique porte sur un tableau qui croise un ensemble de 303 élèves et 30 variables binaires. Il est effectué à l'aide du logiciel CHIC (Couturier, 2008). Il porte ainsi sur 30 variables composites constituées en premier d'éléments de la première ligne et en second d'éléments des lignes suivantes. Par exemple, la variable qui est codée FREc, dans le graphe ci-dessous signifie « Français réussite + écouter » soit en clair *: pour réussir en français, il faut écouter*.

Rappelons la signification du seuil de .95 : toute règle implicative, représentée par un arc du graphe reliant deux variables, admet un nombre de contre-exemples inférieur à celui qu'aurait donné le hasard avec la probabilité .05 dans une hypothèse d'indépendance entre ces deux variables. Ainsi des chemins d'association particulièrement intéressante se dessinent :

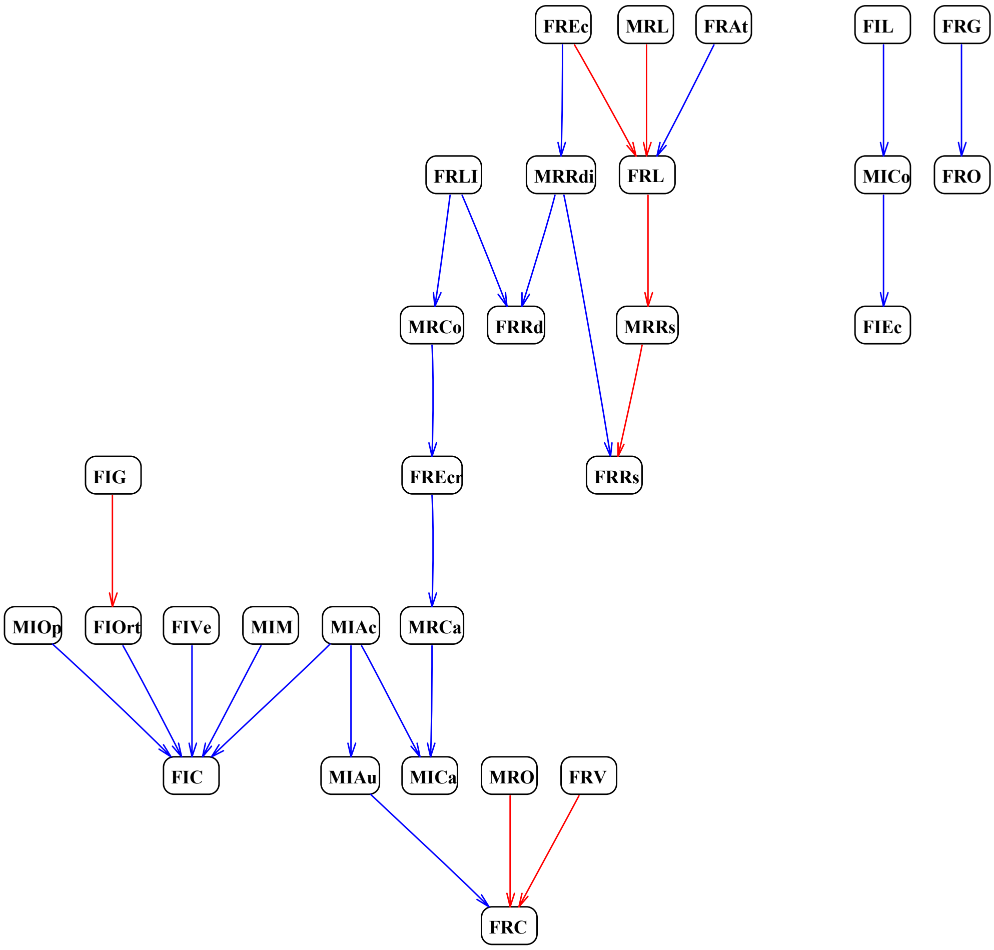


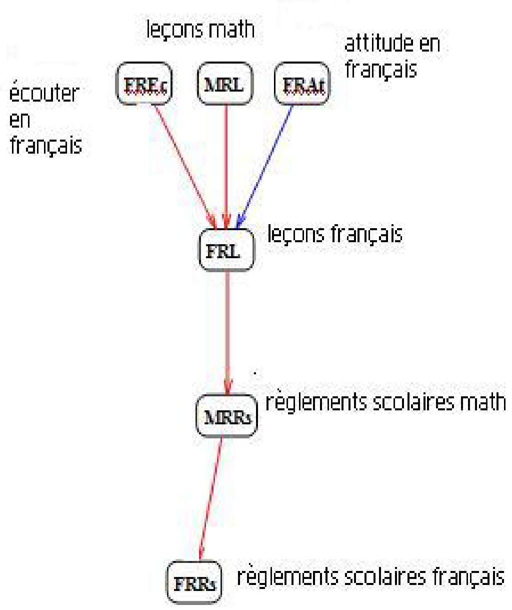
FIGURE 1 - *Graphe implicatif des variables principales au seuil de .95.*

Chemin 1 : MRL⇒FRL⇒MRRs⇒FRRs[13] : ou encore, pour réussir en Mathématiques [bien savoir] [ses] leçons est essentiel⇒pour réussir en Français, [bien savoir] [ses] Leçons. ⇒ [Respecter] des Règles scolaires en Mathématiques ⇒ pour réussir en Français il faut [Respecter] des Règles scolaires. Dans les deux disciplines, [savoir ses leçons] constitue une des composantes suffisantes (au sens logique) du respect des règles scolaires.

[13] Remarquons que FREc, l'écoute en Français, participe aussi de ce chemin.

Chemin 2 : FRLI⇒MRCo⇒FREcr⇒MRCa⇒MICa : ou encore, Pour réussir en Français, [bien] Lire est essentiel ⇒ en Mathématiques [bien savoir] compter ⇒ en Français [bien savoir] Écrire ⇒en Mathématiques [bien savoir] calculer ⇒ ce que l'on apprend en Mathématiques d'important, c'est à [bien] calculer.

Ces chemins sont révélateurs de deux modes de réponses différentes



<u>Dans le premier cas</u>, les élèves considèrent que ce sont des règles scolaires et non disciplinaires qui leur assurent la réussite, quelle que soit la discipline envisagée. Nous proposons de nommer ce chemin celui des « Valeurs scolaires ». Il est révélateur d'une position adoptée par un assez grand nombre d'élèves qui est soit une façon de répondre aisément aux questions posées[14], soit une façon de concevoir la réussite. En tous les cas, la hiérarchie est particulièrement intéressante, car elle révèle une différence disciplinaire, malgré l'apparente similarité d'inféodation de la réussite aux valeurs scolaires.

FIGURE 2– *Chemin 1*

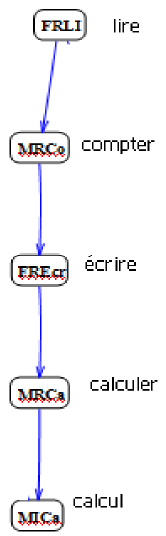
En effet, les valeurs scolaires identifiées comme nécessaires à la réussite le sont d'abord en Français, puis en Mathématiques. De même, les leçons en Français dominent l'écoute – celle-ci contribuant au bon apprentissage des leçons – tandis qu'en Mathématiques les leçons sont identifiées par un plus petit nombre d'élèves, en rapport peut-être avec les volumes horaires respectifs. Notons l'inclusion de « leçons » dans « règlements scolaires » reproduisant l'extension qui fait passer du maître à l'école, i.e. du spécifique au général.

Les trois préceptes de la réussite en Français: Écouter en classe, Avoir une bonne attitude, Apprendre ses leçons, sont donc évoqués dans les réponses d'un groupe important d'élèves. Ces énoncés nous disent la façon dont les élèves reconstruisent un règlement, dans lequel celui qui est le plus retenu est sans doute celui de l'apprentissage des leçons, la règle d'or en quelque sorte. A cette règle d'or sont associés les deux autres principes, mais pratiquement de façon dichotomique. Nous pourrions avancer l'hypothèse que les élèves retiennent ou que les parents et les enseignants tiennent deux discours légèrement différents : « Pour réussir en Français, il faut apprendre ses leçons et bien écouter en classe" d'une part, d'autre part "Pour réussir en Français, il faut apprendre ses leçons et se tenir bien en classe ». Le sens de l'implication serait expliqué

[14] Les questionnaires sont longs. Cependant, les questions concernant la réussite et ce que l'on a appris d'important ont été posées à des moments différents.

par l'ordre des préceptes concernant les leçons et la conduite en classe dans les discours tenus.

Pour ce chemin toujours, les variables supplémentaires qui en sont typiques ou qui y contribuent le plus, à un seuil de risque inférieur à 0,15 sont : des classes particulières, B et E, des établissements, Chopin et Toulouse-Lautrec, soit des milieux favorisés à très favorisés où l'on sait (pense savoir) que la réussite tient au respect de certaines règles de l'institution[15]. La variable la plus typique/la plus contributive est une classe (donc un maître) particulière : ce serait donc un effet maître que celui de concevoir les réussites disciplinaires selon le respect de valeurs scolaires.



Dans le second chemin sont liées des *faire* emblématiques de l'école républicaine : lire, écrire, compter/calculer, faire rattachés aux deux disciplines non moins emblématiques. **C'est pourquoi nous nommons** ce chemin « Piliers de l'école » (les fondamentaux ?). Cette cohérence est aussi révélatrice des proximités des modes de réponse, et, comme nous le supposons, des façons de prendre conscience des espaces disciplinaires. Ici, ce sont donc par des pratiques ou des activités que ces dernières sont identifiées.

Notons que les *techniques* « écrire » et « calculer » tirent bénéfices respectifs du « lire » et du « compter » sans en être totalement assujettis. Par exemple, on peut écrire son nom avant de savoir le lire.

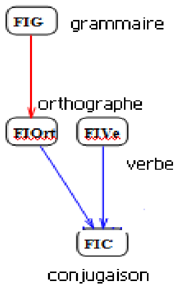
FIGURE 3 – *Chemin 2*

Le sens des relations implicatives peut être entendu de plusieurs façons : soit le « bien lire » et « bien compter » sont considérés comme davantage « acquis » par ces élèves de CM1 et CM2, et à ce titre, participant moins de la réussite disciplinaire ; soit ces positions indiquent que certains d'entre eux comprennent que le savoir lire en particulier est au contraire encore à acquérir car impliquant la compréhension et non pas la seule technicité comme l'écriture. Cette fois, les variables les plus typiques, ou qui contribuent le plus à ce chemin sont l'inscription dans la pédagogie Freinet, l'appartenance à cet établissement. Cela ne nous surprend guère, tant les postulats des enseignants se revendiquant de ce mouvement pédagogique et les pratiques dans les classes (Reuter dir., 2007b) insistent sur les activités des élèves comme principe organisateur des espaces disciplinaires. Certes le fait d'appartenir à un milieu très défavorisé l'est aussi, mais dans une moindre mesure[16], mais sans doute attentif aux éléments fondamentaux du « socle » des disciplines-clés.

[15] Voir en particulier les travaux de l'équipe ESCOL. Mais nous attirons l'attention sur le fait qu'il s'agit ici de déclarations, et non de pratiques. L'écoute nous semble être une pratique scolaire différente selon les pédagogies.

[16] Cet établissement, entièrement inscrit dans un mode de pédagogie de type Freinet est situé dans

Enfin*, un dernier chemin* sera retenu ici : FIG⇒ FIOrt ⇒FIC. Il est révélateur de la cohésion entre la grammaire, l'orthographe, la conjugaison comme sous-domaines de la discipline Français. Il est aussi, en tant que hiérarchie, révélateur du rôle prépondérant de la conjugaison : en effet, les discours des élèves ne cessent de faire référence à ces exercices de conjugaison qui semblent envahir l'espace de la classe de Français, ou à ces « tableaux de conjugaison » qui semblent être l'exact correspondant des tables de multiplication pour beaucoup (Lahanier-Reuter, Reuter, ARCD).



La grammaire apparaît ainsi comme génératrice et régulatrice des compétences orthographiques et a fortiori de celles liées aux règles de conjugaison. Ces sujétions successives, bien restituées par l'A.S.I., émergent de façon d'autant plus remarquable que ce sont de jeunes enfants qui les expriment. D'autres raisons causales peuvent intervenir, comme l'usage pour l'orthographe, mais celles-là, bien connues, sont tôt ressenties à l'école (seuil d'implication 0.99).

Notons aussi que le concept « verbe » est moins associé à l'action ou la description d'un état mais aux règles plus prégnantes de la conjugaison, celle-ci se situant au confluent de l'orthographe et du verbe ; loi et point d'application se coordonnent pour donner vie et sens à la conjugaison.

FIGURE 4 – *Chemin 3*

2.4 Les sujets « atypiques »

Ces chemins ont été choisis non seulement pour leur intérêt pour la recherche évoquée, mais aussi parce que chacun d'eux illustre une forme particulière des recours aux identifications des sujets atypiques.

2.4.1 Ce que nous avons appris d'important en Français : La grammaire, l'orthographe et la conjugaison

Nous nous intéressons ici aux sujets qui contribuent de façon exceptionnelle à ce chemin, c'est-à-dire qui sont plus que les autres élèves responsables de son élection. Ils sont 17 (soit 6% des élèves interrogés). C'est ici que nous voyons en œuvre la dualité des deux espaces variables-élèves. Partant d'un chemin, nous revenons vers l'ensemble des élèves et identifions ceux dont la distance à ce chemin, conceptuellement interprété, est minimale. Parmi les variables supplémentaires qui les lient, nous retrouvons l'effet maître : 7 d'entre eux proviennent de la même classe[17] et quatre des classes ne sont pas représentées.

---

une banlieue « très défavorisée » de Lille. Mais ce n'est pas le seul établissement de notre corpus dans ce cas.

[17] Il s'agit d'une classe de CM2 de la banlieue parisienne, en milieu « très favorisé », qui fonctionne en pédagogie « traditionnelle ».

Mais nous disposons aussi bien sûr des réponses des élèves aux autres questions du questionnaire, et nous avons relevé tout d'abord les particularités de leurs réponses à tout ce qui a trait à la discipline Français.

Il vient que ces élèves ont, au travers des questionnaires (rappelons qu'ils ont été distribués sur une période de temps importante), une attitude pérenne, le Français se décline invariablement au travers de ces trois sous-domaines

*« *Le français sert à apprendre l'orthographe, la grammaire, la conjugaison* » (Faustine, Toulouse Lautrec, CM2 à la question « selon toi, à quoi sert le Français ? »)

* « quand je fais du français je fais: de la conjugaison, de la grammaire, de l'orthographe, de la lecture, des dictées » (Chloé, Courbevoie, CM2) à la question « Qu'est ce que tu fais en classe de français ?»

ou deux des sous domaines « *On apprend les tables de conjugaison, dictées, orthographe, on écrit des textes, on passe des brevets* » (Alicia, Hélène Boucher, CM1),

ou sous un seul, celui de la conjugaison : « *Le français sert à apprendre la conjugaison* » (Alice, Sorlin, CM1) à la question « A quoi sert le Français ? ».

Si nous considérons à présent les réponses de ces mêmes élèves aux questions portant sur les autres matières, nous constatons qu'elles ne se déclinent pas selon des sous-domaines. D'autre part, même si nous pouvons associer certains contenus (Lahanier-Reuter, Reuter, 2011), nous distinguons des thèmes supplémentaires abordés qui ne sont pas congruents avec ceux qu'évoque le Français : « *Les mathématiques servent à calculer vite, à apprendre les tables de multiplications* (cf les règles de la grammaire et des conjugaisons) et *aussi à s'en sortir dans la vie* » (Priscilla, Daudet, CM2), qui évoque (mais elle n'est pas la seule) le monde quotidien d'usage des contenus disciplinaires mathématiques, disons leur utilité sociale. On peut apprendre des actions spécifiques « *A placer les pays sur une carte* » en Histoire-Géographie, « [à faire] *des opérations, des problèmes* » en Mathématiques (Coralie, Sorlin, CM2). Ainsi, le traitement de la discipline Français est spécifique à ces élèves.

Nous en retenons que ces découpages en sous-disciplines ou sous-domaines de la discipline Français, qui la caractérisent pratiquement (références) peuvent être des facteurs de construction des formes de conscience disciplinaire des élèves.

Remarque

Procédons comme nous le faisons pour déterminer la variable supplémentaire contribuant la plus au groupe optimal des élèves (cf. Gras et al 2009) relativement à ce chemin. Nous obtenons alors, avec une loi binomiale de paramètres 27 (nombre d'élèves de la classe) et 17/303, fréquence des élèves du groupe optimal sur l'ensemble total des élèves, la probabilité que le seul hasard conduise à plus de 6 élèves de cette classe de CM2 dans ce groupe optimal est :

$$\mathbf{Prob[Z > 6] = = 1 - \sum_{k=0}^{k=6} C_{29}^{k}\left(\frac{17}{303}\right)^{k}\left(\frac{303-17}{303}\right)^{29-k}}$$

soit par une approximation de Poisson Prob[Z>6] # 0,03. Ainsi, le nombre d'élèves de cette classe de CM2 n'avait qu'une chance sur 30 d'atteindre ce nombre 7 si la répartition des 303 élèves était faite au hasard. Cela renforce la responsabilité de cette classe dans l'élaboration du chemin et donc la signification que nous avons exprimée. Cette exceptionnalité des 7 élèves concernés nous autorise l'expression « atypiques » que nous utilisons ici.

Dans le § 1-3, nous avons proposé l'objectif d'extraire les sujets « atypiques » relativement au chemin FIG⇒ FIOrt ⇒FIC *(grammaire-orthographe-conjugaison)* en traitant le tableau de données initial auquel on aurait adjoint la variable numérique des typicalités des élèves relativement à ce chemin.

Considérant donc cette variable comme variable-intervalle, nous choisissons d'en rechercher une partition optimale (maximisant la variance inter-intervalle) en 5 intervalles. A l'aide de CHIC, menu intervalle, nous obtenons la partition des typicalités suivante :

Ty1 de 0 à 0.116 ; Ty2 de 0.497 à 0.514 ; Ty3 de 0.533 à 0.533 ;

Ty4 de 0.61 à 0.684 ; Ty5 de 0.995 à 0.995.

On observe alors que l'intervalle Ty5 implique transitivement la chaîne FIG⇒FIOrt⇒FIC et que les deux variables supplémentaires qui confortent cette implication sont justement la classe de CM2, évoquée ci-dessus dans le 2-4-1, dite de milieu social Très favorisé. Ce qui corrobore, ici par une autre approche statistique, ce qui a été dit plus haut et permet de désigner les élèves « atypiques » positifs.

Les élèves « atypiques » négatifs, élèves étant donc le plus en désaccord avec cette forme de conscience disciplinaire exprimée par le chemin FIG⇒FIOrt⇒FIC, sont ceux dont la typicalité constitue l'intervalle Ty1. Ils sont en nombre 7 dont 4 appartiennent à un milieu social très défavorisé.

2.4.2 2.4.2 Lire, écrire, compter, ce qu'il faut maîtriser pour réussir

Ils sont 12 élèves seulement dont la typicalité ou la contribution est exceptionnelle relativement au chemin FRLI⇒MRCo⇒FREcr.

Là encore, l'une des variables supplémentaires choisies est une variable explicative : sur ces 12 élèves, 7 fréquentent un établissement d'un quartier « socialement très défavorisé ».

Si nous explorons, comme précédemment, les réponses de ces élèves aux différentes questions, nous enregistrons une permanence, à la fois en mathématiques et en français, des thèmes « lire » en français, « compter » - parfois « calculer »- en mathématiques, écrire » en français. Les Mathématiques servent « A savoir compter comme par exemple quand on va au magasin si on ne sait pas compter on ne peut pas dépenser d'argent » (Célia, Daudet CM2) ou « *Ça peut servir plus tard à savoir calculer* (Mégane, Verhaeren, CM2). Le français sert « *à connaître mieux sa langue et les mots compliqués. On apprend à écrire, à lire, à conjuguer les verbes* » (Clément, Toulouse-Lautrec, CM2). Le Français est présent dans la vie à l'extérieur de l'école « *pour écrire des lettres aux amis* » (Victoria, Pasteur, CM1), de même que les Mathématiques qui sont « *partout, les insectes on peut les compter, les fleurs* » (Mégane, Verhaeren, CM2).

Mais compter quoi ? Quelquefois des unités, quelquefois des opérations[18] puisqu'il faut faire attention à « *ne pas me tromper dans mes opérations, bien mettre les retenues* » (Sarah, Verhaeren, CM2). Écrire ou lire quoi ? Quelquefois il semble que ce soit des textes, le Français peut servir « *A mieux nous comprendre puis pour plus tard pour notre métier il faut écrire des lettes et si on veut faire un métier comme journaliste*

[18] Dans ce cas, compter serait employé pour calculer.

il ne faut pas avoir de fautes, puis savoir lire pour lire les lettres de notre chef » (Franck, Daudet, CM2) ; quelquefois des mots isolés : « *ce que l'on apprend d'important en Français, c'est à écrire. Les différents mots que je connais pas et aussi les verbes et les terminaisons* » (Alexandre, Chopin, CM2). Ces élèves se caractérisent aussi par le fait que ces faire, lire, écrire et compter soient des faire très généraux, dont ils semblent ne pas cerner exactement les contours. En revanche, leurs identifications de ces deux espaces disciplinaires reposent toutes les deux sur des activités, c'est-à-dire se traduisant par des actions.

Le respect des règles scolaires apparaît donc comme facteur de la réussite disciplinaire et gage d'une meilleure intégration sociale.

Quelle interprétation en termes disciplinaires des *« atypiques » négatifs du chemin des comportements scolaires ?*

Nous nous intéressons au dernier chemin relevé, et plus particulièrement encore, à la règle : MRRs⇒FRRs ([Respecter] Règles scolaires en Mathématiques⇒ Règles scolaires en Français). En effet, il est frappant de constater à quel point la plupart des élèves ont adopté des stratégies de réponses identiques pour ces deux disciplines. Par conséquent nous proposons cette fois de nous pencher sur les élèves qui dérogent à cette règle implicative, soit, dans le cadre de cette communication, aux « atypiques » négatifs.

Nous en avons relevé dix, qui sont tous, sauf un, des CM2. Leurs répartitions selon les classes, les milieux socioculturels laissent supposer un léger effet du maître, mais ne permettent pas de conclure à une catégorisation sociale. Certes, ils fréquentent tous des établissements qui ne revendiquent pas de mode pédagogique concerté, mais leur faible nombre est insuffisant pour conclure à une nette influence de ce facteur. Nous le retenons plutôt comme une hypothèse à vérifier.

En revanche, l'examen de leurs réponses aux différentes questions portant tour à tour sur la discipline Français et sur celle des Mathématiques révèle une particularité des réponses de ces élèves : les réponses aux questions successives portant sur le Français montrent une cohérence importante, tandis que celles sur les Mathématiques apparaissent davantage comme indépendantes les unes des autres.

Expliquons-nous. Ces élèves, lorsqu'il leur est demandé « Qu'est-ce que tu as appris d'important en Français ? » puis « Qu'est-ce que tu fais en cours de Français ? » et enfin (bien plus tard) « A quoi fais-tu attention en cours de Français », répondent en désignant un (ou des) contenu(s) disciplinaire(s), qu'ils reprennent dans la seconde question en l'(les) étendant, et dans la troisième en le(s) précisant parfois :

« [ce que l'] *On apprend* [d'important en Français, c'est] à *bien écrire sans faire de fautes*, à bien savoir parler et à savoir *conjuguer* », « [en cours de français] j'apprends de la grammaire, j'apprends à *conjuguer*, à lire et à *bien écrire sans fautes* » et « je fais attention à être propre, *à bien écrire*, faire des phrases complètes et à *ne pas faire de fautes* » (Flora, CM2)

En revanche, ces mêmes élèves, aux mêmes questions déclinées cette fois en Mathématiques, n'adoptent pas cette « règle ». Les contenus qu'ils citent ne sont pas identiques :

« [ce que l'on apprend d'important en Mathématiques] *les fractions, la division* », « [en cours de Mathématiques] quand je fais des maths, je fais du calcul: les puissances de 10, les multiplications, des soustractions, des additions et des divisions. On écrit soit en nombre soit en lettre », et « [je fais attention] *à tout* ». Autrement dit, les critères favorables à la réussite en français sont plutôt les *savoir-faire,* en leur attribuant une valeur modale (ou fait bien, on fait des fautes…)*,* alors qu'en mathématique ce sont plutôt les savoirs que l'on sait ou que l'on ne sait pas.

3 Conclusion

Partant du croisement sujets x variables et de la prise en compte de variables supplémentaires (descripteurs objectifs), nous avons rappelé comment il était possible en A.S.I. de structurer l'ensemble des sujets selon d'autres critères (subjectifs), signatures de comportements, d'attitudes, d'opinions, etc.. Ces critères sont choisis à partir d'une des représentations des variables (arbre des similarités, arbre cohésitif, graphe implicatif) sur la base de l'interprétation de l'expert d'un élément de la représentation choisie, en termes conceptuels. Cette interprétation conduit à l'expression valuée de typicalité ou de contribution de chaque sujet et de chaque variable supplémentaire à l'élection de cet élément graphique, par exemple un sous-graphe de règles, voire une règle seule. Les valeurs de typicalité et de contribution permettent alors de définir de façon duale une structure métrique sur l'ensemble des sujets au moyen d'une distance de typicalité ou de contribution.

A travers une situation scolaire où l'on cherche à identifier chez de jeunes élèves leur conscience disciplinaire, on examine d'une part la structuration des variables à partir d'un questionnaire ouvert portant sur l'attitude des élèves à l'égard du français et des mathématiques et, d'autre part, nous appuyant sur la dualité des deux espaces, le cas des élèves extrêmes ou atypiques quant à cette attitude. Ainsi, l'étude de ces élèves atypiques nous permet de mettre en évidence une tendance, celle de la dissociation entre les degrés de conscience disciplinaire dans les deux disciplines « majeures » de l'école primaire, d'origine didactique probable comme on le rencontre bien souvent dans le secondaire. Cette mise en alerte nous autorise dans ce cas à envisager, dans nos futures investigations, et éventuellement dans une intervention didactique, des attentions plus spécifiques à ces différences.

Références

[1] Clément, M. (2005). *La conscience disciplinaire des élèves de sixième. Étude comparée dans deux collèges de milieux socioculturels différents.* Mémoire de Master 2, Sciences de l'éducation, Université de Lille 3.

[2] Cohen-Azria, C., Hassan, R., et Tutiaux-Guillon, N. (2011) Quelles représentations des disciplines français, sciences et histoire-géographie chez des élèves de cycle 3 ?, *Deuxième Colloque International ARCD*, Villeneuve d'Ascq, 20-22 janvier 2011.

[3] Couturier, R. (2008). Statistical implicative analysis. In CHIC : Cohesive Hierarchical Implicative Classification, *Studies in Computational Intelligence*, Volume 127. Berlin-Heidelberg : Springer Verlag, 41–52.

[4] Diday, E. (1972) *Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes*. Thèse d'Etat, Université de Paris VI.

[5] Gras, R., Ag Almouloud, S., Bailleul, M., Larher, A., Polo, M., Ratsimba-Rajohn, H. et Totohasina, A. (1996). *L'implication Statistique. Nouvelle méthode exploratoire de données.* Grenoble : La Pensée Sauvage.

[6] Gras, R., Kuntz, P., et Briand, H. (2001a). Les fondements de l'analyse statistique implicative et leurs prolongements pour la fouille de données, *Mathématiques et Sciences Humaines* n° 154-155, 9-29

[7] Gras, R., Diday, E., Kuntz, P. et Couturier, R. (2001b). Variables sur intervalles et variables-intervalles en analyse statistique implicative, *Actes VIIIèmes Rencontres de la S.F.C.,* Université de Pointe-à-Pitre.

[8] Gras, R., et Kuntz, P. (2008). An overview of the Statistical Implicative. In R. Gras, E. Suzuki, F. Guillet and F. Spagnolo (Eds) *Statistical Implicative Analysis* Springer-Verlag, Berlin-Heidelberg, 11-40

[9] Gras, R., et Régnier, J.C. (2009). Origine et développement de l'Analyse Statistique Implicative. In R. Gras, J.C. Régnier & F. Guillet (Eds). *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités* Toulouse : Ed. Cépaduès. 6-16

[10] Gras, R., et Couturier, R. (2011). Spécificités de l'Analyse Statistique Implicative (A.S.I.) par rapport à d'autres mesures de qualité de règles d'association. In J.C. Régnier, R. Gras, F. Spagnolo & B. Di Paola(Eds) *Quaderni di Ricerca in Didattica - GRIM* (ISSN on-line 1592-4424) Université de Palerme, 19-57.

[11] Kus, A. (2006). *Les élèves de CM2 et leur conscience disciplinaire en Arts Plastiques*. Mémoire de Master 1, Sciences de l'éducation, Université de Lille 3.

[12] Lahanier-Reuter, D. (2001). Un algorithme de regroupements de modalités de variables en analyse implicative de données. *Mathématiques et Sciences Humaines*, 47-59.

[13] Lahanier-Reuter, D. (2010). Différences affichées et différences masquées entre élèves dans une classe de mathématiques: l'apport de l'analyse didactique. *Actes du IIè congrès International de didactiques.* Girona.

[14] Lahanier-Reuter, D. (2011). Étudier les différenciations des élèves au travers des pratiques de classes et des pratiques langagières des enseignants In J.C. Régnier, R. Gras, F. Spagnolo & B. Di Paola(Eds) *Quaderni di Ricerca in Didattica - GRIM* (ISSN on-line 1592-4424) Université de Palerme, p.283-299

[15] Lebart, L., Piron, M. et Morineau, A. (2006). *Statistique exploratoire multi-dimensionnelle.* 4ème éd. Paris : Science sup, Dunod.

[16] Lécuyer, C. (2005). *La conscience disciplinaire des élèves. Les élèves décrivent-ils différemment en français et en technologie ?* Mémoire de Master 1, Sciences de l'éducation, Université de Lille 3.

[17] Lerman, I.C. (1981). *Classification et analyse ordinale des données.* Paris : Dunod

[18] Marhem, A. (2006). *La conscience disciplinaire en éducation musicale chez les élèves de CM2.* Mémoire de Master 1, Sciences de l'éducation, Université de Lille3.

[19] Polo, M. (1996). *Le repère cartésien dans les systèmes français et italien : étude didactique et application de méthodes d'analyse statistique multidimensionnelle.* Thèse de l'Université de Rennes 1.

[20] Reuter, Y. (2007). La conscience disciplinaire, présentation d'un concept. *Éducation et Didactique*, Vol 1, n°2, 55-70.

[21] Reuter, Y. (2011). Les catégories de contenus et leurs modes d'organisation selon les élèves de fin d'école primaire : comparaison entre Français et Mathématiques. *Deuxième Colloque International ARCD*. Villeneuve d'Ascq, 20-22 janvier 2011.

[22] Ségismont, C. (2006). *La conscience disciplinaire des élèves de CM2. Étude complémentaire,* Mémoire de Master 2, Sciences de l'éducation, Université de Lille 3.

[23] Sève, L. (2005). *Émergence, complexité et dialectique.* Paris : Odile Jacob

[24] Van Meenen, E. (2006). *La conscience disciplinaire des élèves de CM2. Étude auprès de deux classes aux modes de fonctionnement pédagogique différents.* Mémoire de Master 2, Sciences de l'éducation, Université de Lille 3.

Annexe

Classification des sujets en fonction de leur contribution à un ensemble de règles

Cet ensemble de règles (implicatives ou de similarité) peut être extrait en totalité ou partiellement d'une classe de la hiérarchie cohésitive (ou de similarités) ou d'un chemin du graphe implicatif ou d'un ensemble connexe de ce graphe, dans le cas de l'implication statistique. Le choix est guidé par la signification de cet ensemble de règles en terme de concept(s) auquel (auxquels) le chercheur a attribué une signification.

Les données sont alors constituées d'une matrice croisant les sujets et les règles retenues non exhaustives. Au croisement ligne-colonne on affecte la valeur de la contribution (ou la typicalité suivant le choix retenu) du sujet à la règle, donnée par CHIC. La distance d_C(**1**) ci-dessus, est retenue. La méthode d'analyse qui conduit à une classification respectant une certaine topologie de l'espace des sujets (proximité) est celle de la carte dite auto-organisée construite par l'algorithme de Kohonen. On obtient une grille généralement rectangulaire (par exemple k = 3x4 cases) qui correspond à un réseau de neurones contenant au final une partition « organisée » de l'espace de sujets ou d'une partie de cet espace. A chaque neurone (une des k cases) on associe un voisinage formé des neurones situés à une distance inférieure à un seuil défini sur la base d'un choix d'un indice de proximité (par exemple la contiguïté). Ce choix permet de disposer d'une classification des sujets et de sous-ensembles de sujets qui seront d'autant plus proches que leurs contributions au(x) concept(s) le seront également. L'algorithme de Kohonen se déroule en un certain nombre d'itérations qui convergent vers la classification attendue (Lebart et al, 2006).

Pour des raisons de difficulté de programmation des calculs liés à cette application, nous nous contenterons ici simplement de cette évocation qui peut servir de piste de recherche à des familiers de l'A.S.I..