



HAL
open science

Évolution vers MARIAMA 2. Gestionnaire d'hypothèses et d'analyse de données

Robert Nicolai

► **To cite this version:**

Robert Nicolai. Évolution vers MARIAMA 2. Gestionnaire d'hypothèses et d'analyse de données. GDRE 1172 du CNRS. 1996. <hal-05316214>

HAL Id: hal-05316214

<https://hal.science/hal-05316214v1>

Submitted on 15 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Evolution vers MARIAMA 2
gestionnaire d'hypothèses et d'analyse de données

Projet de développement
des fonctions de "gestionnaire d'hypothèses" et "d'aide à l'analyse de données"
en vue de leur application
à la gestion des systèmes classificatoires en général.

1996

R. NICOLAI

Le logiciel MARIAMA "gestionnaire d'hypothèses" version 1 avait été conçu dans un premier temps pour aider à la recherche comparative, et dans un deuxième temps comme un auxiliaire à l'élaboration de dictionnaires. Il s'applique sur une base de données qu'il exploite par les fonctions "classiques" : recherches, tris, importations, exportations, habituellement attendues d'un programme de gestion de fiches ; mais aussi par des fonctions particulières en rapport avec sa finalité propre.

Dans sa version 2, certaines des caractéristiques essentielles de sa structure et de ses fonctions particulières seront généralisées pour créer un nouvel outil d'aide à la constitution de tout projet classificatoire qui porte sur des systèmes d'objets.

Cette évolution de MARIAMA permettra une utilisation généralisée du logiciel pour la gestion de projets, l'analyse et le développement d'hypothèses de travail dans d'autres secteurs des sciences humaines que ceux de la linguistique, dans le domaine des sciences de la vie ou tout simplement, pour certaines analyses documentaires.

On trouvera dans ce document la description des concepts généraux et des principales structures utilisés pour cette nouvelle version.

Notion d'objet et de description dans MARIAMA 2

Sont considérées comme des '*objets*' des unités organisées en systèmes, identifiables par des caractéristiques différentielles partiellement ou complètement stables.

Les objets peuvent être *structurellement simples* ou *complexes* dans leur définition (cf. en linguistique, les unités élémentaires, éventuellement caractérisées sur deux plans : expression et signification). Dans ce dernier cas, les structurations et organisations formelles de gestion qui seront proposées devront être "clivées" en relation avec le degré de complexité attribué à ces objets. Le clivage le plus simple est obtenu par un dédoublement de certaines rubriques, lorsque les limites des objets coïncident sur les différents plans ; lorsque ce n'est pas le cas, la notion de 'plans d'analyse' (cf. p. 11) peut remplacer le clivage.

Sont considérées comme des '*descriptions*' des structures minimales formées de deux ensembles d'objets, tels que les éléments de l'un soient identifiés comme des *descripteurs* et que les éléments de l'autre soient identifiés comme des *unités simples*, par rapport à une pertinence donnée, et que nous ayons une application de l'ensemble des unités simples dans l'ensemble des descripteurs (on parlera de *synthèse*) ; la relation inverse établie dans ce cadre est une *analyse*, elle sera formellement représentée par une matrice.

Il est posé en principe que la classification d'objets est basée sur de telles *descriptions* et que le nombre de descriptions affectées à un objet n'est pas limité.

La description d'un objet (ou ensemble d'objets) donné est :

- . soit une synthèse, et dans ce cas l'objet est défini de façon élémentaire comme appartenant à une classe d'objets du même type
- . soit une analyse, et dans ce cas, il est défini de façon non élémentaire par l'ensemble des relations inverses qui renvoient à la même image dans l'application.

Un objet (ou ensemble d'objets) donné est donc à la fois:

- . potentiellement assignable n fois à n classes d'objets du même type ; les classes étant éventuellement ordonnées (synthèse)
- . potentiellement analysable n fois par rapport à n classes de n éléments constituants élémentaires organisés selon des règles particulières ; les classes étant éventuellement ordonnées (analyse).

On remarquera, bien évidemment, que la relation de description est toujours la même, mais la définition des objets en tant que résultat d'analyse ou de synthèse est une affaire de décision par rapport aux domaines de définition de l'application.

D'un point de vue "pragmatique", c'est une "décision théorique externe" qui établira la distinction entre ce qui est 'élémentaire' et donné à priori, et ce qui est construit et 'descriptible'.

Le terme "n fois" signifie *n applications de l'opération en question sur le même objet*, ce qui, plus concrètement, renvoie pour le premier cas à la prise en compte de différents systèmes de référence, et donc de pertinence (cf. : p. 11, la notion de plans d'analyse), et pour le deuxième cas à la prise en compte d'une structuration de l'objet en "constituants non-élémentaires" sur lesquels l'opération s'appliquera.

Les structurations de base dans MARIAMA 2

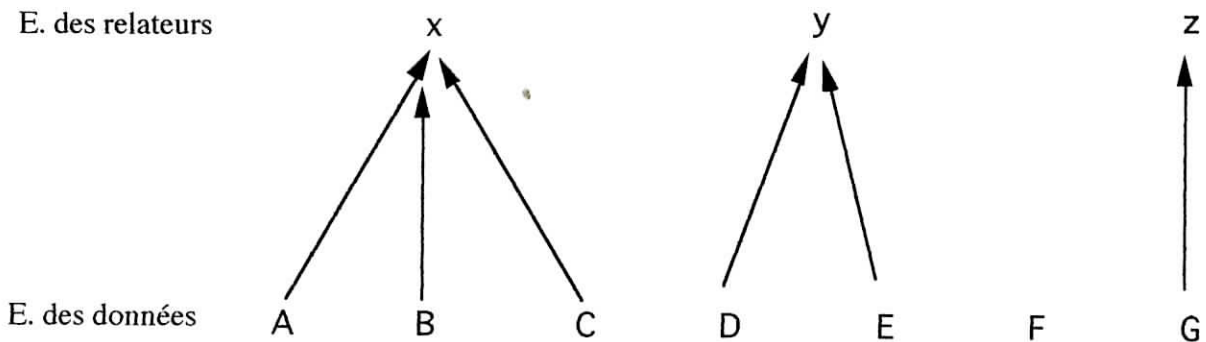
MARIAMA est organisée autour de la notion générale de *hiérarchie*. Sont considérées comme des '*hiérarchies*' des structures composées par des ensembles d'objets d'un niveau donné, liés entre eux par la relation qu'ils entretiennent avec un relateur ou un spécificateur qui se définit à un autre niveau. C'est la *relation* entre ces niveaux qui établit la hiérarchie. On dira que l'ensemble des *relateurs* ou des *spécificateurs* est situé au niveau supérieur de la relation et que l'ensemble des *objets* reliés ou spécifiés est situé au niveau inférieur.

En rapport avec la notion de description précédemment établie, deux modalités d'organisation hiérarchique sont concernées :

- Celle qui conduit à la constitution d'hypothèses sur les objets.

Largement mise en évidence dans la version 1, elle est en rapport avec la notion de *relateur* liant plusieurs objets. Une *hypothèse* est une relation qu'on établit entre plusieurs objets appartenant à un même type de données, elle génère une classification ; celle-ci peut être unique ou non.

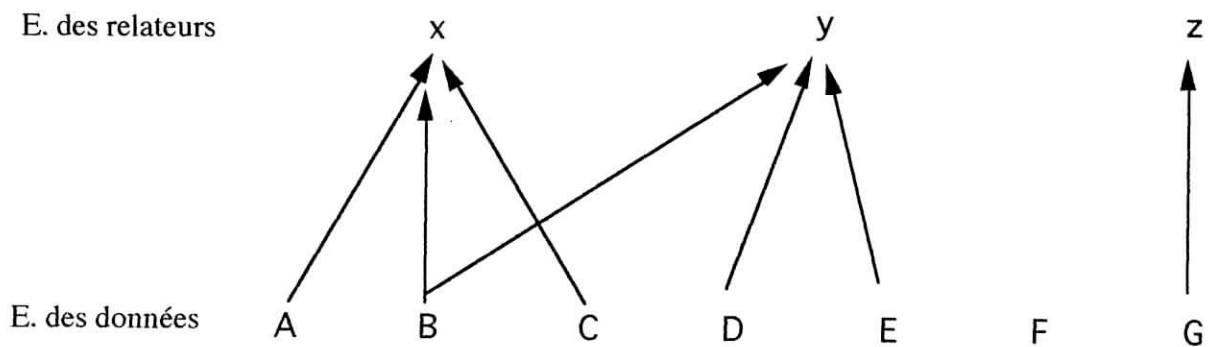
Lorsque la relation est unique, elle correspond formellement à une application de l'ensemble des données dans l'ensemble des relateurs:



Application de l'ensemble des données dans l'ensemble des relateurs

mais ce cas n'est qu'un cas particulier de l'ensemble des relations possibles.

Dans le cas général, nous avons affaire à des relations non spécifiées pour lesquelles un même objet peut éventuellement avoir plusieurs "images" dans l'ensemble des relateurs.



Cas général : relations non spécifiées

Dans ce dernier cas, un objet de niveau inférieur est en relation non pas avec une seule unité relateur, mais avec plusieurs ; on a alors affaire à des structures potentiellement "ambigües" ou "indéterminées". On peut cependant s'attendre, dans certains cas et après analyse des données, à ce que la structure générale soit décomposable en applications distinctes de l'ensemble des données dans l'ensemble des relateurs, définies par rapport à différentes pertinences (cf. la notion de description), et donc à ce que l'indétermination ou l'ambigüité soit levée. Ainsi, la figure précédente pourrait éventuellement correspondre à deux applications de l'ensemble des données dans l'ensemble des relateurs, définies par rapport à deux pertinences différentes :

pertinence a : {A,B,C} {D,E} {F} {G}

pertinence b : {A,C} {B,D,E} {F} {G}

La définition des pertinences relève, bien entendu, d'une analyse indépendante et de choix théoriques externes.

- Celle qui conduit à l'analyse interne des objets.

Une analyse spécifie un objet par rapport à un système de traits . Les traits spécifiants qu'elle utilise sont en nombre fini, en principe préexistants, indépendants et établis a priori. Ils appartiennent à un autre niveau que celui des données et sont éventuellement ordonnés. L'ensemble des spécifications attribuées à un objet dans une analyse particulière peut être assimilé à une 'matrice' à n lignes et n colonnes.

a1	a2
b1	b2
...	...
x1	x2

Un même objet peut être défini par n matrices de ce type, correspondant soit à une description, éventuellement hiérarchisée, de ses éléments composants (cf. un mot qui peut être décrit comme une unité simple, ou qui peut être décomposé en une succession de sons ou de caractères) soit à une présentation, éventuellement hiérarchisée, de plusieurs descriptions alternatives.

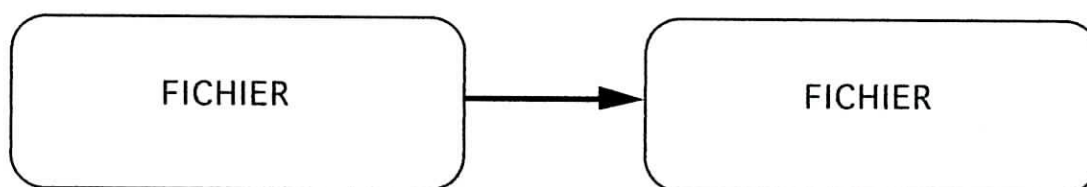
Un système approprié de *spécificateurs* permet d'introduire des relations d'ordre et/ou une organisation différenciée entre les différentes analyses ou hypothèses qui concernent un objet donné ou des ensembles d'objets.

Description physique

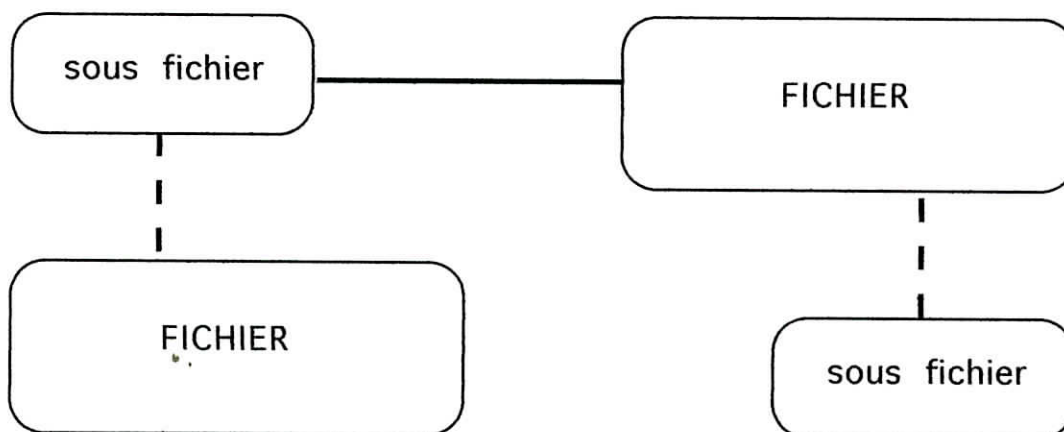
1. La structure générale des descriptions

L'organisation des descriptions est la même, qu'il s'agisse de synthèses ou d'analyses. Il s'agit d'une structure composée de deux fichiers liés. Ce sont les affectations données aux éléments de cette structure qui introduiront son utilisation comme outil de gestion d'analyses ou comme outil de gestion d'hypothèses.

Schéma de la structure minimale de description



L'élaboration de cette structure dans MARIAMA implique pour chacun des fichiers, l'adjonction d'un sous fichier ou d'un fichier lié supplémentaire. Le lien principal sera orienté dans un sens ou dans l'autre selon que l'on a affaire à une gestion d'analyses ou d'hypothèses.



- . Pour la gestion d'analyses, il sera orienté du fichier vers le sous fichier
- . Pour la gestion d'hypothèses, il sera orienté du sous fichier vers le fichier

2. La structure de gestion d'hypothèses

La structure élémentaire de gestion d'hypothèses est une description ainsi constituée :

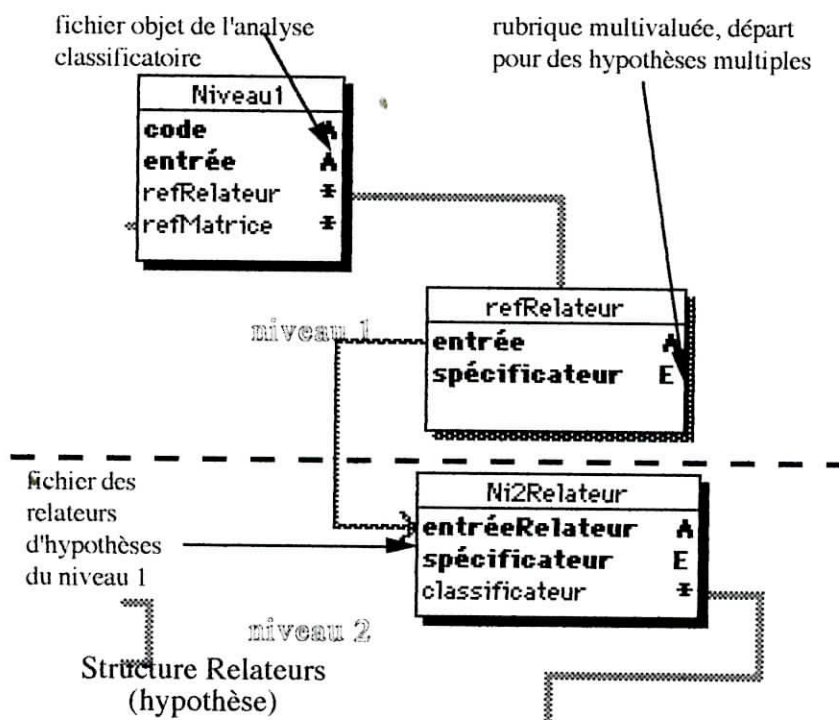
- le fichier des 'objets', nommé fichier Niveau 1
- le fichier des 'relateurs', nommé fichier Niv2Relateur

tels que le fichier Niveau 1 pointe sur le fichier Niv2Relateur.

Cette structure élémentaire modélise une application du fichier Niveau 1 sur le fichier Niv2Relateur, elle n'autorise donc pas la création d'hypothèses indéterminées ou ambiguës. Pour permettre cela, la structure doit être complexifiée par l'adjonction soit d'un fichier lié supplémentaire, soit d'un sous-fichier interne aux fiches de Niveau 1.

Ce nouveau fichier ou sous fichier, désigné par refRelateur dans le schéma, fonctionne alors comme une rubrique multivaluée et permet de caractériser la fiche à laquelle il est lié par différentes valeurs prises par le relateur. Une même fiche peut ainsi être liée à plusieurs relateurs ; on a donc ainsi le départ pour l'élaboration d'hypothèses multiples et la généralisation souhaitée est obtenue.

Schéma de la structure de gestion d'hypothèses:



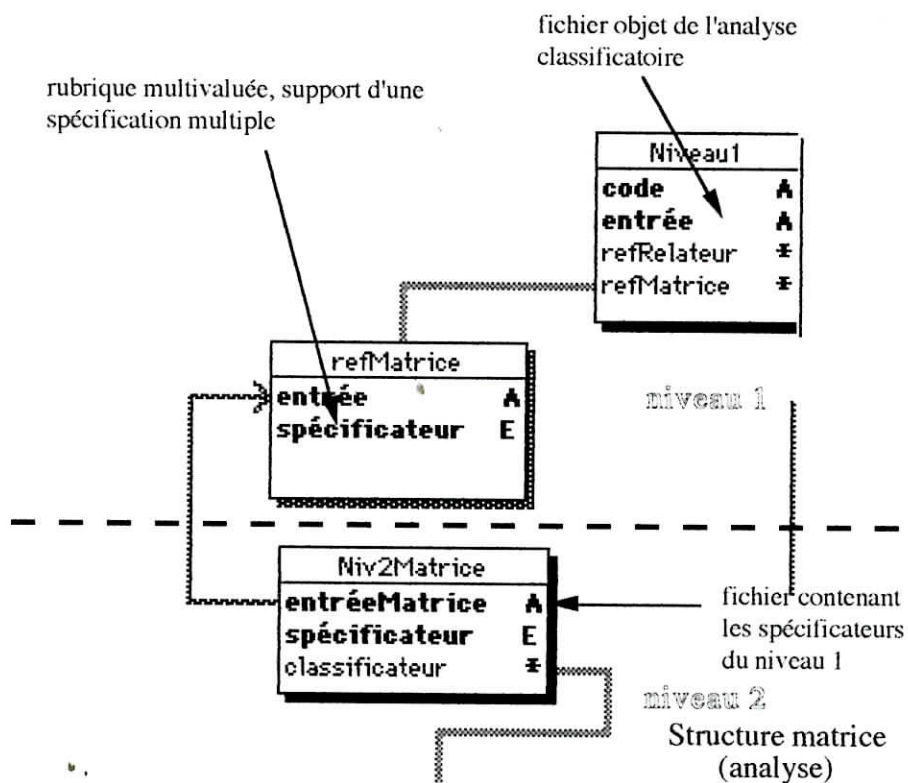
3. La structure de gestion d'analyses

La structure élémentaire de gestion des analyses est une description ainsi constituée :

- le fichier des 'objets', nommé fichier Niveau 1
 - le fichier des 'traits', nommé fichier Niv2Matrice
- tels que le fichier Niv2Matrice pointe sur le fichier Niveau 1.

Cette structure modélise une application du fichier Niv2Matrice sur le fichier Niveau 1, elle n'autorise donc pas des spécifications indéterminées ou ambiguës. La possibilité d'attribuer n analyses à une fiche particulière est réalisée lorsque l'on définit la rubrique qui reçoit le lien comme une rubrique multivaluée, soit en la transformant en sous fichier, soit en en faisant le point d'arrivée d'un fichier lié.

Schéma de la structure de gestion d'analyses:

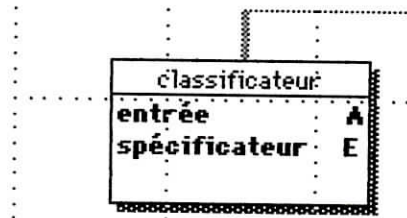


Intégration d'une relation d'ordre.

Les relations d'ordre peuvent concerner aussi bien le niveau des données que celui des traits et des relateurs.

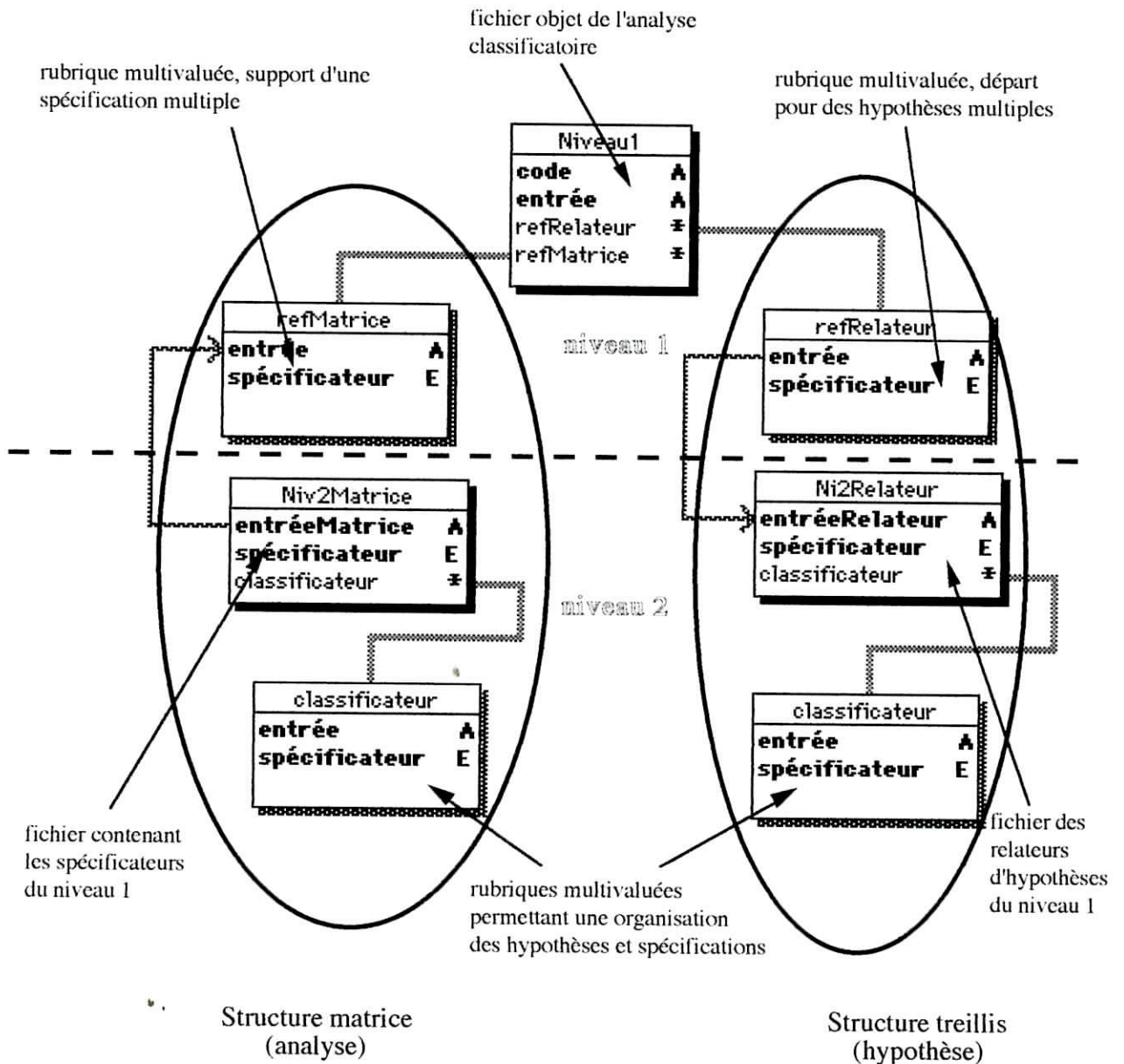
Ainsi, dans RefMatrice et refRelateur, les rubriques "spécificateur" permettent d'introduire un ordre sur les valeurs affectées aux rubriques 'entrée' du niveau 1.

Corrélativement, le fichier 'classificateur' est lié aux fichiers du niveau 2 sur lesquels il peut être établi et une organisation particulière (rubrique entrée), et une autre relation d'ordre (rubrique spécificateur) .



Organisation classificatoire générale

Le schéma ci-dessous présente l'ensemble fonctionnel composé par les structures de gestion d'analyses et d'hypothèses munies de leurs spécificateurs. Il correspond à la structure minimale et supposée fondamentale, proposée pour la gestion des systèmes classificatoires en général.



Cette organisation classificatoire est le 'coeur' du gestionnaire. Selon les besoins nécessités par un projet classificatoire particulier, elle sera intégrée dans une structure plus adaptée comprenant différentes rubriques finalisées utiles pour le projet . Elle pourra également être *stratifiée*, c'est à dire suivie ou précédée, en amont et/ou en aval, de couches structurelles supplémentaires (cf. niveau 0 et niveau 3 de MARIAMA vers. 1.0). Elle pourra enfin, être reprise en parallèle dans son propre niveau (cf. les '*plans d'analyse*' de MARIAMA vers. 1.0) ou dans les couches structurelles supplémentaires car l'un des intérêts de MARIAMA est bien de construire non pas une mais plusieurs structurations classificatoires indépendantes sur les mêmes données.

Ces structurations peuvent alors être mises en relation. Autrement dit, on peut simultanément faire des regroupements de données sur la base de pertinences diverses, puis croiser ou ne pas croiser les résultats, les modifier éventuellement les uns par rapport aux autres au fur et à mesure que la recherche se développe.

On retient donc deux notions descriptives utiles :

- plans d'analyse : chaque structuration classificatoire indépendante correspond à un '*plan d'analyse*'. Le nombre des plans d'analyse est une caractéristique structurelle du logiciel. La version standard de MARIAMA Vers. 1.0 (par exemple) permet de travailler sur cinq plans d'analyse différents.
- stratification : l'introduction de couches supplémentaires est justifiée par rapport à une utilisation particulière. Ainsi, dans la version standard, utilisée à des fins de recherche linguistique, il y a une couche structurelle intitulée '*niveau 0*', qui contient les données de référence, auxquelles il est toujours possible de revenir. Ce niveau est lui même composé par une structure hiérarchisée particulièrement fonctionnelle pour le travail de construction de dictionnaires.

Affectations des rubriques

Niveau 1

code	code unique de la fiche
entrée	référence de l'objet traité
refRelateur	origine du sous-fichier supportant le lien concernant les hypothèses
refMatrice	origine du sous-fichier supportant le lien concernant les analyses

refRelateur

entrée	référence du lien d'hypothèse
spécificateur	rubrique susceptible d'introduire une relation d'ordre dans les entrées

refMatrice

entrée	référence du lien d'analyse
spécificateur	rubrique susceptible d'introduire une relation d'ordre dans les entrées

Niv2Relateur

entréeRelateur	référence de l'hypothèse
spécificateur	rubrique susceptible d'introduire une relation d'ordre dans les entrées
classificateur	origine du sous-fichier supplémentaire supportant le lien concernant l'ordre ou la multivalence des hypothèses

Niv2Matrice

entréeMatrice	référence de l'analyse
spécificateur	rubrique susceptible d'introduire une relation d'ordre dans les entrées
classificateur	origine du sous-fichier supplémentaire supportant le lien concernant l'ordre ou la multivalence des spécifications

classificateur

entrée	référence du classificateur
spécificateur	rubrique susceptible d'introduire une relation d'ordre dans la spécification des entrées du Niveau 2

Exigences et fonctions dans MARIAMA

La fonction essentielle de MARIAMA n'est pas d'être un simple serveur de base de données, même sophistiqué, bien qu'il soit évident que cette fonction est nécessaire. Sa fonction essentielle est davantage d'être un outil d'aide à l'élaboration de projets classificatoires.

Elle doit donc permettre de structurer, et de restructurer en fonction des résultats obtenus, les données de la base qui lui est liée. En conséquence, les données dans MARIAMA ne sont jamais "intouchables", elles sont modifiables continuellement en rapport avec l'état des recherches en cours. L'outil ainsi constitué a cette fonction de

- donner les moyens de procéder aux recherches susceptibles d'entraîner des modifications sur les données,
- faciliter ces procès de modification,
- établir l'ensemble des "hypothèses" et "indexations" qui en dépendent.

Les exigences et fonctions fondamentales pour ce type d'utilisation sont répertoriées ci-dessous.

- Utilisation d'espaces de travail inclus dans les fiches.

L'espace fiche doit comprendre deux parties : celui qui reçoit les informations concernant les données proprement dites et un espace supplémentaire prévu pour fonctionner en tant qu'espace "de travail" pour les besoins ponctuels de l'utilisateur. Dans les rubriques appropriées de cet espace peuvent être introduites des indexations, des commentaires et des annotations diverses. Bien évidemment elle peuvent faire l'objet de recherches, tris, etc. , comme toutes les autres rubriques.

- Accès aux sources des données.

Accès toujours immédiat au niveau le plus bas qui est celui du recueil des informations "brutes", c'est à dire avant tout traitement ou interprétation induite par l'utilisation du gestionnaire, et cela, quel que soit le niveau hiérarchique auquel l'on se situe.

- Perspectives et navigation.

La notion de perspective renvoie aux différents modes d'exploration (c'est à dire de navigation inter-fichiers utilisant — et n'utilisant que — la structure des liens de la hiérarchie) des données. Une sélection peut être consultée directement dans le fichier auquel elle appartient, mais elle peut aussi être appréhendée à d'autres niveaux, en utilisant cette structure des liens.

- **Gestion identique à tous les niveaux.**

Toutes les types d'opérations possibles à un niveau donné de MARIAMA sont également possibles aux autres niveaux de ses hiérarchies... sauf si cela entraîne une contradiction logique.

- **Conservation des sélections.**

Le travail effectué avec MARIAMA peut impliquer d'importants mouvements de navigation à l'intérieur d'un fichier et entre fichiers. Il est donc nécessaire de pouvoir travailler avec les différentes sélections effectuées, qui parfois résultent de recherches complexes. Indépendamment de la sélection courante, doit être conservée:

- . la sélection précédente
- . la sélection établie au départ de la séquence de travail
- . une autre sélection optionnelle supplémentaire qui permet de conserver une relation d'ordre, c'est à dire un tri particulier de la sélection en question.

- **Opérations ensemblistes.**

Possibilité de conserver dans des "ensembles" des sélections particulières que l'on souhaite pouvoir retrouver ultérieurement et d'utiliser les entités ainsi construites dans un calcul ensembliste.

- **Opérateurs de recherche.**

Ne sont mentionnés ici que les opérateurs complexes.

- recherche par saturation : à partir d'une sélection initiale de fiches, une recherche extensive est appliquée sur l'ensemble du fichier ; elle porte sur les valeurs contenues dans une rubrique discriminante préalablement sélectionnée. Le résultat contiendra la réunion de tous les résultats de recherche concernant chacune des valeurs contenues dans la rubrique discriminante pour la sélection considérée.

- recherche par enchaînements : opérateur qui introduit une recherche extensive effectuée sur une rubrique à partir d'un fichier-outil construit par l'utilisateur en fonction de ses besoins spécifiques. Ce fichier, intitulé "*famille*", permet de créer et de lier entre-eux un certain nombre "*d'indicateurs de recherche*" dont la mise en rapport est jugée pertinente pour le projet de l'utilisateur. Il est constitué en tant que ressource autonome et peut être soit l'amorce d'une base de connaissances, soit celle d'un stock d'hypothèses indépendantes, utilisables à toutes fins utiles et modifiables selon les besoins. En tout état de cause, il fonctionne comme une macrocommande de recherche multicritère limité à la rubrique choisie.

- recherche sur appariements: opérateur qui, à l'intérieur d'une sélection donnée, recherche l'ensemble des fiches liées par une relation formelle de synonymie ou d'homonymie et ne conserve que celui là.

La relation de synonymie ou d'homonymie peut être réduite à une relation de quasi-homonymie ou de quasi-synonymie, en ne faisant porter la recherche que sur le début de la rubrique.

- recherche sur les valeurs distinctes d'une sélection: opérateur qui introduit une recherche consistant à présenter l'inventaire des valeurs concernées par une rubrique discriminante.

- **"Travail" en miroir.**

Il s'agit d'un plan d'analyse particulier qui a une fonction originale car il doit permettre l'élaboration d'un dossier critique construit à partir de l'ensemble des analyses et hypothèses préalablement établies sur les données, et proposer une organisation alternative à celle qui définit cet ensemble initial d'hypothèses et d'analyses, sans pour autant modifier les informations de référence.

Tout se passe comme si ce fichier était un "fichier image" du niveau 2, mais modifiable en fonction de l'analyse critique que l'on fait des informations qu'il contient. On peut ainsi :

- utiliser telles quelles toutes les informations de niveau 2 contenues dans la base
- les modifier dans le nouveau fichier, si on le juge utile, mais sans que cela ait pour autant des répercussions dans les données d'origine.

Cela permet à une personne extérieure d'utiliser les résultats de la base de façon indépendante.

Ce plan d'analyse particulier autorise aussi, dans un espace approprié, l'écriture directe d'un article et le report de toutes les sources et références utiles pour un tel travail. Son accès fait l'objet d'une autorisation particulière.

- **Opérations de modification.**

Elles concernent aussi bien les fichiers de données que ceux d'hypothèses.

- modification des systèmes d'hypothèses : par définition, une hypothèse est susceptible d'être confirmée... ou infirmée. En conséquence le gestionnaire se donne les moyens de modifier, supprimer ou recomposer les hypothèses construites.

- dédoublément des fiches de données : lors de leur saisie, il se peut que des entrées aient été considérées comme des unités simples alors qu'une analyse ultérieure conduit à souhaiter qu'elles soient scindées en unités indépendantes. Une fonctionnalité dite de "dédoublément", permet de résoudre ces problèmes en effectuant les recodages nécessaires.

- *Fonction de remplacement de chaîne* : cette fonction est proche de celle que l'on trouve dans un logiciel de traitement de texte mais ici elle est sélective en ce sens que l'on choisit le fichier et la rubrique sur laquelle elle doit s'appliquer. En conséquence, et plus particulièrement en ce qui concerne la transcription des formes, elle permet de procéder à des modifications de caractères qui peuvent être justifiées pour se mettre, par exemple, en conformité avec de nouveaux choix de normes.

- *Fonction de "nettoyage" de rubrique* : elle permet tout simplement de supprimer ou d'ajouter un nombre donné de caractères (une chaîne) en début ou en fin de rubrique. Sa fonction est donc différente de la précédente : il ne s'agit plus de modifier une chaîne particulière quelle que soit sa place dans le champ ; il s'agit, au contraire, quelle que soit la chaîne, de la modifier si elle appartient à la zone finale ou initiale de la rubrique.

- modification continue des rubriques par report généralisé : pour une sélection donnée, aussi importante fût elle, il faut pouvoir reporter en une seule opération une information dans une rubrique, quelle qu'elle soit. Si la rubrique en question se trouve être liée à une autre rubrique, cela doit être pris en compte dans le procès de modification.

- *Gestion de la hiérarchie des niveaux et fonctions spécifiques.*

De par la nature de MARIAMA, il est évident qu'un ensemble de fonctions spécialement adapté à gérer ses hiérarchies est nécessaire.

- Balancements de fichiers :

On identifiera ainsi une fonction générale de "*balancement de fichier*". On dira qu'on balance un fichier A sur un fichier B lorsque l'on effectue un report global des fiches d'une sélection du fichier A dans le fichier B en créant les fiches appropriées et les codages nécessaires au report dont il s'agit.

Techniquement, les balancements de fichiers peuvent avoir lieu aussi bien entre fichiers de niveaux différents qu'entre des fichiers de même niveau. Les deux types de balancement ne répondent pas aux mêmes contraintes et c'est à l'utilisateur d'évaluer l'opportunité de procéder ou non à l'opération.

Le balancement est l'une des modalités qui permettent de reporter les fiches liées d'une sélection donnée dans un nouveau fichier. Bien évidemment cette modalité est indépendante et d'une nature toute différente de celle que matérialise la construction des hypothèses. Au plan pratique, elle présuppose que les codes établissant le lien entre les deux fichiers aient été intégrés dans le fichier de départ. MARIAMA possède des fonctions pour permettre cette intégration.

- Mise à jour des fichiers :

A côté de cette fonction de balancement de fichier il y a celle de "*mise à jour de fiches*". Dans ce dernier cas il ne s'agit pas de créer de nouvelles fiches, mais uniquement de "mettre à jour" un fichier avec les informations contenues dans un autre ; il s'agit donc d'une fonction de modification.

- Report de rubrique :

Il s'agit, pour une sélection donnée, de la possibilité de reporter les valeurs contenues dans une rubrique vers une autre rubrique du même fichier ou d'un fichier lié.

Exemples complémentaires:

Trois exemples simples d'hypothèses portant sur des données linguistiques :

Le linguiste qui procède à un travail comparatif — que ce soit à fin d'établir des apparentements génétiques, de faire une analyse dialectologique, phonologique, sémantique, morphologique ou autre — propose des hypothèses sur les données qu'il compare.

I. *Classification arborescente simple*

Forme générale . Soit l'ensemble $\{A, B, C, D, C\}$ qui correspond à des objets soumis à classification.

Une division peut être établie à partir de rapprochements dépendants des choix de pertinence et des outils d'analyse de l'utilisateur, soit, par exemple:

$X = \{A, B, C\}$ et $Y = \{D, C\}$. Elle renvoie à une partition de l'ensemble et implique donc une indépendance entre les ensembles segmentés. chaque élément regroupé n'appartient qu'à un et un seul ensemble.

Toutefois, la possibilité d'identifier les ensembles segmentés par plus d'une forme (et donc de se donner les moyens de "gérer une indétermination" à ce niveau, doit être prise en compte.

Exemple linguistique . Le rapprochement ci dessous effectué entre les formes suivantes appartenant à des dialectes du songhay, langue parlée dans la Boucle du Niger:

Dialecte	Entrée	Signification
kaado	<u>debe</u>	damer, tasser en frappant
touareg	<u>atAbbi</u>	taper avec la paume de la main
bozo	<u>tEbE</u>	piétiner

.....

.....

1. établit une *hypothèse sur les relations qu'elles entretiennent*
2. s'appuie sur une *pertinence historique* qui définit le mode d'existence de la relation.
3. se définit par rapport à un type de *classification arborescente simple*.

..

II. *classification arborescente en relateurs*

Forme générale . Soit l'ensemble $\{A, B, C, D, C\}$ qui correspond à des objets soumis à classification.

La même division que celle proposée précédemment peut être établie soit : $X = \{A, B, C\}$ et $Y = \{D, C\}$.

Mais pour des raisons particulières l'on peut être amené à envisager que plusieurs segmentations de l'ensemble soient simultanément possibles, et donc que certains objets puissent appartenir à plusieurs regroupements à la fois . Soit donc le rapprochement $Z = \{B, D\}$ qui a pour conséquence que les éléments /B/ et /D/ appartiennent à deux ensembles à la fois. Cette opération là pouvant être théoriquement continuée.

Dans ces conditions, il est évident que la classification arborescente simple correspond à un cas particulier de celle ici présentée. Soit donc, le cas où aucun élément n'est référé à plus d'un ensemble. En conséquence on ne distinguera plus entre les deux modalités de classification.

Exemple linguistique . Au plan de la reconstruction interne des formes, la possibilité de pouvoir reconstruire un objet avec une forme ou avec une autre suivant que l'on intègre ou non certaines langues dans les comparaisons , illustre le problème.

III. *classification matricielle*

Forme générale . Soit l'ensemble $\{A, B, C, D, C\}$ qui correspond à des objet soumis à classification.

L'étude des données conduit à proposer une description de chacun de ces objets à partir d'un système de traits indépendants qui a la fonction d'analyseur. Chaque objet est ainsi défini par une matrice de traits appartenant au système analyseur. la structure de la matrice est éventuellement hiérarchisée.

Exemple linguistique . L'exemple le plus approprié est celui d'une analyse en traits phonétiques ou sémantiques (analyse componentielle).

Une hypothèse n'a de sens que dans un but de validation ou d'invalidation. Cette validation relève d'une pertinence externe. Elle se fait par l'utilisation classique des outils théoriques, méthodologiques et instrumentaux dont dispose l'utilisateur, et par rapport aux critères de pertinence qu'il s'est donné.

Autrement dit, par exemple, l'on peut simultanément faire des regroupements de données sur la base de considérations sémantiques, sur la base de considérations phonétiques, sur la base de considérations "historiques" etc. , puis croiser ou ne pas croiser ces résultats, les modifier éventuellement les uns par rapport aux autres au fur et à mesure que la recherche avance.

On peut organiser ces regroupements par rapport à un même type de classification ou par rapport à plusieurs.

Pratiquement, MARIAMA permet de *travailler sur cinq plans d'analyse différents*. Ces cinq plans d'analyse correspondent à cinq systèmes d'hypothèses différents qui permettent tous d'établir des relations sur les données de base. Des modalités particulières de fonctionnement ont été attribuées par défaut à chacun d'eux : elles les prédéterminent pour certaines fonctions. En rapport avec ses besoins et ses objectifs propres, l'utilisateur leur attribuera des fonctions spécifiques. Ces plans d'analyse particuliers constituent ce que l'on appellera le "niveau2" de MARIAMA.