

MODÉLISATION DU COMPORTEMENT DES AGENTS ET DE LA DYNAMIQUE RÉSIDEN TIELLE

Si le secret du fonctionnement du vivant se trouve dans sa structure, le secret de sa structure se trouve dans la logique qui a présidé à son élaboration.

J. Piaget

Dans les chapitres précédents, nous avons mis en évidence les déterminants de la mobilité résidentielle. Nous avons également vu que les agents résidentiels (les ménages) sont caractérisés par des propriétés que nous avons organisées en six variables (statut d'occupation, le revenu, le niveau d'études, l'âge, la composition du ménage et les réseaux de relations sociales). Pour chaque agent, la combinaison de ces différentes variables agit sur le comportement résidentiel à chaque étape de la vie. ***Comment formaliser la structuration de ces variables afin de proposer un modèle permettant d'inférer le comportement résidentiel des agents ?*** Deux types d'approches permettent de répondre à cette interrogation. D'une part, les méthodes statistiques car elles sont précisément conçues pour permettre le passage de l'observation à la loi. Et d'autre part, les méthodes issues de l'intelligence artificielle, parce que leur vocation est de permettre aux ordinateurs de traiter des connaissances probabiliste, logique ou factuelle. La combinaison de ces deux approches est possible grâce au formalisme des réseaux bayésiens.

En effet, les réseaux d'inférence bayésiens sont le résultat d'une convergence entre ces deux disciplines et constituent l'un des formalismes les plus complets et cohérents pour l'acquisition, la représentation et l'utilisation des connaissances par des ordinateurs [Naïm, et al., 2004 ; Cornuéjols et al., 2002 ; Xiang, *op. cit.*, 2002] mais aussi pour le

calcul de probabilités conditionnelles. Dans le cadre de notre recherche, les réseaux sont utilisés pour modéliser la mémoire décisionnelle des ménages quant à ce qui concerne leur comportement de mobilité.

Après avoir présenté les cadres théoriques et méthodologiques de la mise en œuvre des réseaux d'inférence bayésiens, nous réalisons dans ce chapitre le modèle comportemental des agents résidentiels avant de dresser de façon détaillée l'algorithme régissant le comportement dynamique des agents durant la simulation.

6.1. Considérations théoriques relatives aux réseaux bayésiens

A la fois outil mathématique de calcul de probabilités conditionnelles et modèle de représentation graphique de connaissances, les réseaux d'inférence bayésiens ou plus simplement réseaux bayésiens doivent leurs noms aux travaux de Thomas Bayes au XVIII^{ème} siècle sur la théorie des probabilités. Ils sont actuellement le résultat des recherches effectuées dans les années 80 par J. Pearl [Pearl, 1986a, 1986b, 1987, 1988] à UCLA (University of California, Los Angeles) et une équipe de recherche danoise de l'université d'Alborg. Encore du domaine de la recherche dans les années 90, ils connaissent maintenant de plus en plus d'applications concrètes par exemple le contrôle de véhicules autonomes ou le diagnostic médical ou encore la fouille de données, la reproduction du raisonnement et du langage, etc.

Les réseaux bayésiens sont des modèles qui permettent de décrire des relations de probabilités conditionnelles entre des faits. Cette représentation repose sur un graphe orienté sans cycle (DAG : *directed acyclic graph*) dans lequel chaque nœud, c'est-à-dire chaque variable du *monde* modélisé, possède une table de probabilité conditionnelle et où chaque arc représente une dépendance directe entre les variables reliées. L'ensemble du réseau représente alors la distribution des probabilités jointes de l'ensemble des variables de manière compacte, en s'appuyant sur les relations d'indépendance conditionnelle entre nœud. A ce titre, les réseaux bayésiens conjuguent deux aspects : une partie qualitative modélisée au moyen d'un graphe et une partie quantitative exprimée par le biais de probabilités (*cf.* Figure 6.1). Les réseaux bayésiens s'appuient donc sur la théorie des graphes et la théorie des probabilités.

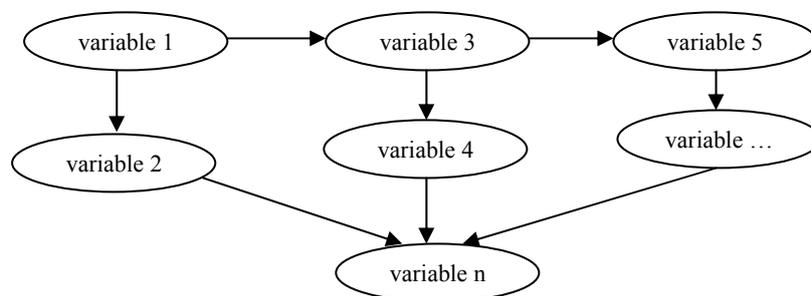


Figure 6.1 : Représentation graphique d'un réseau bayésien à n variables

6.1.1. Les graphes orientés sans cycle

L'idée de base de la théorie des graphes est de proposer un outil de manipulation et d'étude d'un ensemble fini d'objets sur lequel est défini une relation binaire quelle que soit la sémantique de cette relation. La théorie des graphes se donne donc pour objectif d'étudier de manière abstraite un type de structure d'ensemble qui ne dépend que d'une relation binaire entre ses éléments. Les graphes peuvent être alors interprétés comme une description des relations entre paires d'éléments. Le caractère abstrait d'une telle description permet à cette théorie d'avoir des champs d'applications extrêmement vastes et variés. Par exemple, elle est souvent utilisée pour l'analyse des réseaux de villes en géographie [Mathis, 2003].

L'objectif de cette section n'est pas de présenter la théorie des graphes dans son ensemble mais simplement de décrire les diverses notions indispensables à une bonne compréhension du modèle décisionnel des agents que nous allons développer.

6.1.1.1. Notion de graphe

Un graphe est un ensemble de couples d'éléments vérifiant une relation donnée. Cet ensemble peut être fini ou non. Mais en l'occurrence nous nous intéressons uniquement aux ensembles finis car le nombre de variables que nous avons identifiées dans le chapitre précédent pour caractériser les agents résidentiels est fini.

Soit $V = \{v_1, v_2, \dots, v_n\}$ un ensemble fini non vide. Un graphe G sur V est défini par la donnée du couple $G = (V, E)$ où $E \subset \{(u, v) \mid u, v \in V \text{ et } u \neq v\}$. V est alors nommé l'ensemble des nœuds de G . E peut être donc considéré, par extension, comme étant la description de la relation entre chaque paire de nœuds. L'écriture $u \neq v$ interdit l'existence d'une relation réflexive. En effet, certaines définitions acceptent (u, u) comme couple dans E ; ce qui n'est pas notre cas parce qu'aucune variable ne sera en relation avec elle-même. Les distinctions fondamentales entre les divers types de graphes dépendent de la nature exacte des éléments de E c'est-à-dire la nature des relations entre chaque paire de nœuds.

6.1.1.2. Notion d'arc

La notion d'orientation est très importante lorsqu'on utilise des modèles abstraits comme les graphes. Lorsque nous considérons par exemple l'ensemble des villes françaises reliées par autoroute, la relation entre chaque paire de villes est clairement symétrique (*double sens*). En revanche, dans le cas d'un automate, ce n'est pas parce que celui-ci peut passer d'un état A à un état B qu'il pourra passer de B à A (les transitions possibles entre A et B peuvent donc être à *sens unique*). Dans ce cas la relation est qualifiée d'*arc* et dans le cas inverse, elle est qualifiée d'*arête*.

Soit un graphe $G = (V, E)$. Pour tout élément (u, v) de E ,

- (u, v) est un arc et noté $(u \rightarrow v)$ si et seulement si $(v, u) \notin E$,
- (u, v) est une arête et notée $(u - v)$ si et seulement si $(v, u) \in E$.

Comme défini, un arc n'apparaît qu'une seule fois dans E pour chaque paire de nœuds (u, v) de V . Les arcs et les arêtes permettent de définir et de manipuler de façon homogène les graphes orientés, et non orientés. Les graphes orientés constituent le socle des modèles bayésiens et de ce fait, sont le type de graphe qui nous intéresse.

6.1.1.3. Les graphes orientés et quelques notions connexes

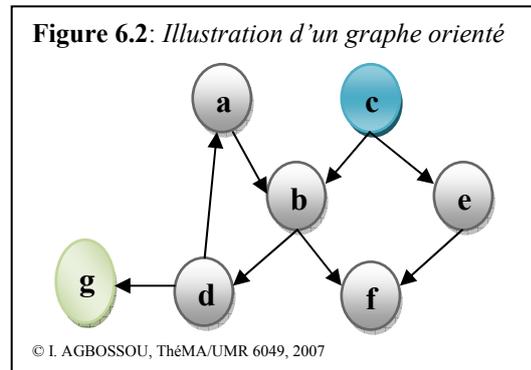
Un graphe $G = (V, E)$ est un graphe orienté noté \vec{G} si et seulement si tous les éléments de E sont des arcs. Ainsi, $\vec{G} = (V, E)$ où $E \subset \{(u, v) \mid u, v \in V, u \neq v \text{ et } (u \rightarrow v)\}$. Pour tout arc $(u \rightarrow v)$ appartenant à E , u est l'*origine* de l'arc et v son *extrémité*. On dit que u est un *parent* (ou *prédécesseur*) de v et v , l'*enfant* (ou *successeur*) de u . Nous noterons Γ_v l'ensemble des parents de v et Ξ_u l'ensemble des enfants de u . De même, on pourra définir l'ensemble des parents ou des enfants d'un sous-ensemble A de V :

- $\Gamma_v = \{u \in V \mid (u \rightarrow v) \in E\}$
- $\Gamma_A = \{u \in V \setminus A \mid \exists v \in A, (u \rightarrow v) \in E\}$
- $\Xi_u = \{v \in V \mid (u \rightarrow v) \in E\}$
- $\Xi_A = \{v \in V \setminus A \mid \exists u \in A, (u \rightarrow v) \in E\}$.

$V \setminus A$ étant l'ensemble V privé des éléments de A .

Dans le graphe \vec{G} , un nœud sans parent est appelé *racine* et un nœud sans enfant est appelé *feuille*. Dans la figure 6.2, nous pouvons avoir les illustrations suivantes :

- $\vec{G} = \{\{a, b, c, d, e, f, g\}, \{(a, b), (b, d), (b, f), (c, b), (c, e), (e, f), (d, a), (d, g)\}\}$
- $\Gamma_b = \{a, c\}$
- $\Xi_b = \{d, f\}$
- $\Gamma_{\{a,b,e\}} = \{d, c\}$
- le nœud c est une racine,
- le nœud g est une feuille.



6.1.2. Les probabilités

Les réseaux bayésiens ont la particularité d'allier deux champs différents des mathématiques dans le but de représenter le manque d'information : la théorie des graphes d'une part, qui fournit le cadre nécessaire pour une modélisation qualitative des connaissances ; et la théorie des probabilités d'autre part, qui permet d'introduire une information quantitative dans les connaissances.

6.1.2.1. Notions de probabilité et de variable aléatoire

La théorie des probabilités propose un cadre mathématique pour représenter quantitativement l'incertain. La présentation que nous allons en faire est forcément tronquée puisqu'elle est orientée vers son utilisation dans le domaine des réseaux bayésiens. En particulier, l'espace sur lequel seront définies les probabilités restera discret et fini. Ce n'est bien sûr pas le cas général mais c'est suffisant pour le développement des modèles que nous allons élaborer.

Soit Ω un ensemble fini non vide, $(\mathcal{E}, \cap, \cup)$ une algèbre c'est-à-dire un triplet offrant la possibilité de faire des calculs sur Ω où \mathcal{E} est un sous-ensemble de l'ensemble des parties de Ω ($\mathcal{E} \subset 2^\Omega$, l'ensemble des parties de Ω ou encore univers). Soit $P: \mathcal{E} \rightarrow [0, 1]$ une fonction à valeurs réelles. P est une probabilité sur (Ω, \mathcal{E}) si et seulement elle vérifie :

- $\forall \mathcal{A} \in \mathcal{E}, 0 \leq P(\mathcal{A}) \leq 1 ;$

- $\forall \mathcal{A}, \mathcal{B} \in \mathcal{E}, [\mathcal{A} \cap \mathcal{B} = \phi] \Rightarrow P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}); \mathcal{A} \text{ et } \mathcal{B} \text{ sont des ensembles mutuellement exclusifs, } \phi \text{ étant l'ensemble vide ;}$
- $P(\Omega) = 1 \text{ et } P(\phi) = 0.$

Tout élément de \mathcal{E} est appelé *évènement élémentaire* sur Ω . Un *évènement* sur Ω est donc une sous-partie de Ω c'est-à-dire une combinaison d'évènements élémentaires. Ω est appelé l'*évènement certain* et ϕ l'*évènement impossible*.

Une variable aléatoire permet de caractériser des évènements (qui sont des sous-ensembles d'évènements élémentaires) par une simple valeur. C'est donc une fonction χ définie sur Ω comme suit :

$$\chi : \begin{cases} \Omega & \rightarrow \mathcal{D}_\chi \\ \omega & \mapsto \chi(\omega) \end{cases}$$

Pour $x \in \mathcal{D}_\chi$, on note alors $\{\chi = x\}$ l'évènement $\{\omega \in \Omega | \chi(\omega) = x\}$, \mathcal{D}_χ étant le domaine de définition de χ . Si le domaine de définition de la variable χ est fini, alors χ est une *variable aléatoire discrète*. Comme nos modèles sont restreints à un Ω fini, les variables seront donc toujours considérées comme discrètes.

Dans la suite, nous allons représenter une variable aléatoire par une lettre majuscule (A, B, ...). La valeur que prend cette variable aléatoire sera notée par la même lettre mais en minuscule (a, b, ...). De même nous noterons P(a) la probabilité que l'évènement A prenne la valeur a ($\{A = a\}$).

6.1.2.2. Notions de probabilité jointe et de probabilité marginale

Une variable aléatoire est donc un moyen pour condenser une information pertinente sur un univers (ensemble des possibilités). Cependant, il faut souvent plus d'une variable aléatoire pour caractériser l'état de l'univers. Nous assimilons ici, l'univers au système cognitif décisionnel de l'agent résidentiel. Il est pratique de décrire ce système grâce à un ensemble de variables permettant de le caractériser à tout moment.

Pour ce faire, il faut lui adjoindre une probabilité sur les différentes variables qui permettent de décrire et de caractériser son état à un moment précis. Cette *probabilité jointe* est construite à partir de la même fonction de probabilité sur $\Omega : P$. La liste des arguments (c'est-à-dire les évènements sur lesquels porte la probabilité) d'une probabilité jointe est donc suffisante pour la caractériser. Ainsi, la probabilité que les évènements A, B, C et D prennent respectivement les valeurs a, b, c et d est notée

$P(a, b, c, d)$. Plus généralement, si U est un ensemble fini non vide de variables aléatoires discrètes ($U = \{X_1, X_2, \dots, X_n\}$) sur Ω , U est appelé le vecteur état du système. Réciproquement, la donnée d'une probabilité jointe d'un ensemble de variables permet de retrouver la probabilité jointe de chacun de ces sous-ensembles. C'est ce qu'on appelle une *probabilité marginale*. L'opération qui permet d'obtenir les probabilités marginales est la fonction *marginalisation* de P sur V (V étant un sous-ensemble de U) définie comme par : $\forall V \subset U, P(V) = \sum_{x \in U \setminus V} P(V, x)$. Dans l'illustration ci-contre,

| $P(t, l)$ | l_1 | l_2 | $P(t)$ |
|-----------|-------------|-------------|--------------|
| t_1 | 0.0578 | 0.0782 | 0.136 |
| t_2 | 0.1604 | 0.0576 | 0.218 |
| t_3 | 0.5118 | 0.1342 | 0.646 |
| $P(l)$ | 0.73 | 0.27 | |

nous mettons en évidence le calcul des probabilités marginales sur les variables aléatoires T et L fictives dont les modalités ou valeurs sont respectivement t_1, t_2, t_3 et l_1, l_2 . Les

valeurs numériques sur fond gris clair représentent les probabilités jointes et celles sur fond gris foncé sont les probabilités marginales. En effet, $P(l_1) = P(L = l_1, T = t_1, T = t_2, T = t_3)$.

6.1.2.3. De la probabilité conditionnelle au théorème de Bayes

Une autre notion fondamentale en calcul des probabilités est celle de *probabilité conditionnelle*. En réalité, pour un évènement ω de l'univers Ω , la valeur $P(\omega)$ est associée au moins implicitement à des conditions de réalisation. Si A et B sont deux variables aléatoires, l'expression d'une probabilité conditionnelle de A par rapport à B se traduit par « étant donné que l'évènement $B = b$ s'est produit, la probabilité que l'évènement $A = a$ se produise (ou se soit produit) est k » et s'écrit $P(a | b) = k$ et se lit « probabilité de a sachant b égale à k ». Cette définition permet d'arriver naturellement au théorème de Bayes [Naïm, et al., 2004, *Op. cit.*] que nous démontrons en annexe 1:

$$P(a, b) = P(a|b).P(b) = P(b|a).P(a) \quad (6.1)$$

En transformant (6.1), on obtient $P(a|b) = \frac{P(b|a).P(a)}{P(b)}$ (Bayes 1)

Plus généralement, $P(a|b, c) = \frac{P(b|a, c).P(a|c)}{P(b|c)}$ (Bayes 2)

Ce théorème est à la base de tout un pan de la statistique nommé, de manière assez compréhensible, la statistique bayésienne. Sans entrer dans les détails, dans l'équation (Bayes 1), le terme $P(a)$ est la probabilité *a priori* de A . Elle est « antérieure » au sens qu'elle précède toute information sur B . $P(a)$ n'est rien d'autre que la probabilité

marginale de A. Le terme $P(a|b)$ est appelée la probabilité *a posteriori* de A sachant B (ou encore de A sous condition B). Elle est « postérieure », au sens qu'elle dépend directement de B. Le terme $P(b|a)$, pour un B connu, est appelée la fonction de *vraisemblance* de A. De même, le terme $P(b)$ est appelé la probabilité marginale ou a priori de B. C'est pourquoi le théorème de Bayes s'écrit aussi :

$$\mathbf{loi\ a\ posteriori\ \cong\ loi\ a\ priori\ \times\ vraisemblance\ \quad (Bayes\ 3).}$$

La statistique bayésienne est donc une approche qui tend à automatiser l'application de la loi a priori sur des quantités inconnues, quitte à effectuer une mise à jour, principalement grâce à l'équation (Bayes 3), lorsque plus de renseignements auront été récoltés.

La généralisation de la loi fondamentale (6.1) permet de calculer la probabilité jointe de plusieurs variables aléatoires à partir de leurs probabilités conditionnelles :

$$P(x_1, x_2, x_3, x_4) = P(x_1|x_2, x_3, x_4) * P(x_2|x_3, x_4) * P(x_3|x_4) * P(x_4) \quad (6.2)$$

D'après l'équation (6.2), on voit que la manipulation des probabilités jointes sur plusieurs variables est une tâche ardue qui implique des algorithmes de complexité exponentielle, en fonction du nombre de variables.

Pour rendre possibles les calculs sur de telles probabilités, on recourt à la notion d'indépendance conditionnelle. En effet, cette notion permet la simplification du second membre de l'équation (6.2). Concrètement, elle consiste à éliminer le calcul des probabilités conditionnelles des variables dont la connaissance n'apporte aucune information supplémentaire. De ce fait, dans le cas d'une application basée sur des données d'enquête, il devient alors possible d'approcher les probabilités marginales par les moyennes statistiques.

P. Frankhauser avait, pour preuve, démontré de façon formelle dans sa thèse de doctorat [Frankhauser, 1991] le passage de la moyenne statistique de données d'observation au calcul de probabilité.

Ainsi, en nous appuyant sur les conclusions de P. Frankhauser, lorsque deux variables X et Y sont conditionnellement indépendantes on a :

$$P(X|Y) \cong \frac{1}{N} \sum_{i=1}^N P_i(X = x|Y = y) * P(Y)$$

En effet, le théorème de Bayes est utilisé dans l'inférence statistique pour mettre à jour ou actualiser les estimations d'une probabilité, à partir des observations et des lois de probabilité de ces observations.

Contrairement à l'école "fréquentiste" qui utilise les propriétés de long terme de la loi des observations sans prendre en considération la variabilité des paramètres, l'école "bayésienne" utilise les probabilités comme moyen de traduire numériquement un degré de connaissance, en l'occurrence approximée par la moyenne statistique (la théorie mathématique des probabilités n'oblige en effet nullement à associer celles-ci à des fréquences, qui n'en représentent qu'une application particulière résultant de la loi des grands nombres). Dans cette optique, le théorème de Bayes peut s'appliquer à toute proposition, quelle que soit la nature des variables et indépendamment de toute considération ontologique.

6.2. Formalisation du modèle comportemental des agents

La mobilité résidentielle est une problématique vaste et complexe. Sa complexité vient du fait qu'il ne s'agit pas d'étudier seulement les trajectoires résidentielles des ménages ou les mutations du parc de logements, mais l'évolution du couple ménages-logements au sein du système résidentiel (*cf.* Figure 2.2).

6.2.1. Revisite du système résidentiel

Dureau et Flores définissent le système résidentiel individuel comme «*la configuration spatio-temporelle définie par l'ensemble des logements habités par un individu au cours de l'année, l'intensité et le rythme de résidence dans chacun d'eux* ». [Dureau, 1999 ; Dureau et Flores 2000]. Reprenant cette définition, nous l'étendons au ménage (agent résidentiel) en élargissant la dimension temporelle à l'espérance de vie de ce dernier ou au nombre d'années passées de façon continue dans la zone géographique considérée.

En clair, nous considérons le système résidentiel du ménage comme étant la configuration spatio-temporelle définie par les logements habités par un ménage au cours de sa vie ou pendant une certaine durée dans un espace géographique bien précis. En l'occurrence, l'espace géographique est délimitée par la zone d'étude. La mobilité spatiale vue comme tout changement de logement sera formalisée dans le cadre de la

mobilité urbaine. Ces déplacements par changement de logements ont des conséquences directes sur la dynamique urbaine. On aurait pu complexifier le système en complétant la figure 2.2 par les acteurs de la politique de l’habitat urbain. La dynamique du système est régie par les diverses interactions entre deux grandes entités : les habitants représentés par les agents résidentiels (ménages) et l’habitat représenté par les logements. Chaque type de ménage est modélisé au travers d’agent cognitif et l’habitat par l’automate cellulaire. La modélisation du volet spatial par automate cellulaire et son couplage avec le SMA est détaillée à la section 6.3. Il s’agit maintenant de voir comment on va exploiter ce système résidentiel pour arriver à une modélisation comportementale des agents résidentiels.

6.2.2. L’inférence comportementale des agents

Un modèle bayésien permet de représenter un ensemble de variables aléatoires pour lesquelles on connaît (ou lorsqu’il est possible de définir la structure du réseau à partir d’un algorithme d’apprentissage) l’ensemble des dépendances. Appliqué aux ménages, ce modèle permet d’inférer de façon dynamique (mise à jour des connaissances d’un ménage à chaque étape de son cycle de vie) le comportement des ménages quant à leur décision de changer de résidence et le choix du type de logement à occuper. La construction d’un tel modèle se décompose en trois étapes distinctes : une représentation qualitative des relations de causalité existantes entre les variables aléatoires, la définition des tables de probabilités jointes puis la spécification numérique des probabilités conditionnelles.

6.2.2.1. Spécification du graphe causal

Dans le quatrième chapitre nous avons présenté l’architecture d’un agent BDI (cf. Figure 4.2). Dans cette représentation, la base de connaissances est le siège de l’inférence cognitive du ménage. Et nous faisons l’hypothèse qu’avant de prendre une quelconque décision de déménagement, le ménage va d’abord évaluer sa satisfaction résidentielle. Le processus aboutissant à une telle évaluation est modélisé par le graphe causal ci-dessous.

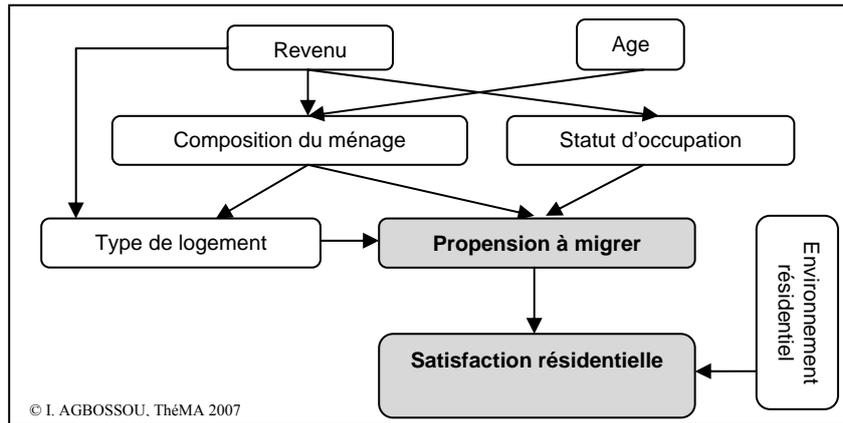


Figure 6.3 : Graphe causal de la base de connaissances du ménage

Ce graphe illustre donc la première étape consistant à modéliser qualitativement le raisonnement d'un ménage (où chaque flèche matérialise un lien de dépendance directe), en amont, quant à une éventuelle possibilité et décision de changer de logement. Il permet également de mettre en évidence les influences entre les variables. La prochaine étape est la définition du *schéma d'inférence* c'est-à-dire les probabilités jointes.

6.2.2.2. Spécification du schéma d'inférence

La spécification probabiliste du réseau d'inférence passe par une représentation utilisable d'une distribution jointe sur l'ensemble des variables. Ainsi, la décomposition de la distribution de probabilité jointe peut se faire en s'appuyant sur l'équation (6.2) :

$$P(\chi_1, \dots, \chi_n) = \prod_{i=1}^n P(\chi_i | \Gamma_{\chi_i}) \text{ où } \Gamma_{\chi_i} \text{ est l'ensemble des parents de } \chi_i. \quad (6.3)$$

Chaque χ_i représente une variable c'est-à-dire chaque nœud du graphe causal de la figure 6.3. Comme le réseau encode la distribution de probabilité jointe pour l'ensemble des variables, il permet donc de calculer n'importe quelle probabilité d'intérêt c'est-à-dire la probabilité de réalisation d'un évènement quelconque. La décomposition est toujours la même et permet de ne spécifier que des probabilités locales, c'est-à-dire les probabilités d'une variable sachant uniquement les variables ayant une influence directe sur elle.

6.2.2.3. Quantification des probabilités conditionnelles

Cette étape consiste à spécifier les tables de probabilité $P(\chi_i | \Gamma_{\chi_i})$ pour chaque variable de la base de connaissances de l'agent. Ceci revient à donner l'ensemble des probabilités de la variable pour chacune de ses valeurs possibles sachant chacune des valeurs de ses parents. Par exemple la variable “*composition du ménage*” possède deux variables parents : *revenu* et *âge*. Ainsi, la quantification de la probabilité conditionnelle $P(\text{composition du ménage} = \text{personne seule} | \text{revenu} = 1000 \text{ €}, \text{âge} > 35 \text{ ans})$ par exemple nécessite non seulement la disponibilité des tables de probabilité des trois variables mais aussi les tables de probabilités de chacune des trois variables conditionnellement aux deux autres.

Ces probabilités peuvent être données par un expert ou bien apprises à partir d'un corpus d'exemples (ce qui nécessite le développement d'algorithmes d'apprentissage du réseau). Dans le cadre de cette recherche, les probabilités marginales sont obtenues à partir des diverses données issues de la revue de littérature sur la mobilité résidentielle notamment les enquêtes logement et les données de recensement de la population. Dans notre cas, les simulations que nous réaliserons au prochain chapitre sont basées sur des données d'enquête. La définition des probabilités conditionnelles conduit ainsi à la réalisation des tables de probabilités jointes. Ces données constituent bien évidemment des informations de base (il s'agit de la connaissance *a priori*) qui seront dynamiquement mise à jour (pour produire la connaissance *a posteriori*) durant chaque simulation grâce aux données complémentaires qui seront fournies en entrées dans le simulateur. D'autres informations relatives à la dimension spatiale sont également requises. Elles font l'objet de la section suivante.

6.3. Formalisation de la dynamique spatiale de l'habitat

Nous modélisons la dynamique spatiale de l'habitat par un automate cellulaire à maille régulière carrée (c'est-à-dire basée sur un carroyage) dont le côté de chaque cellule sera fonction de la résolution spatiale de la zone d'étude. Les règles de transition temporelle sont spécifiées par des chaînes de Markov contraintes.

6.3.1. Spécifications partielles de l'automate cellulaire

Le développement d'un outil de simulation de type multi-agents ou automate cellulaire demande préalablement une démarche conceptuelle, de formalisation indépendante de l'informatique, qui doit si possible s'appuyer sur un corpus théorique [Langlois, 2006]. Nous avons répondu à cette exigence scientifique au travers du cinquième chapitre relatif à la modélisation des automates cellulaires. L'objectif de cette section est donc de procéder à la formalisation du modèle d'automate cellulaire applicable à la thématique de la mobilité résidentielle. Nous définissons donc les spécifications formelles relatives aux notions de champ cellulaire, d'état, de voisinage et configuration initiale.

- *Le champ ou le réseau cellulaire* : c'est l'ensemble des cellules disposées dans l'espace selon un arrangement topologique bien défini, qui correspond en l'occurrence à une grille de carroyage dont la résolution spatiale (que nous noterons *ResolutionSpatiale*) sera définie par l'utilisateur du simulateur. La résolution spatiale est la structure de données qui encapsule les dimensions de chaque cellule. Chaque cellule ayant une forme géométrique carrée, la résolution spatiale définit donc le côté d'une cellule. Le réseau cellulaire peut être carré ou rectangulaire, composé de cellules carrées. Il correspond à la cartographie de la zone d'étude. Le nombre total de cellules (que nous notons *NbreTotalCellules*) composant le champ cellulaire est donc obtenu par la relation suivante :

$$NbreTotalCellules = \frac{LargeurChampCellulaire * HauteurChampCellulaire}{ResolutionSpatiale^2} \quad (6.4)$$

où *LargeurChampCellulaire* et *HauteurChampCellulaire* désignent les dimensions physiques de la zone d'étude. En effet, *NbreTotalCellules* n'est rien d'autre que le produit du nombre de lignes ($NbreLignes = \frac{HauteurChampCellulaire}{ResolutionSpatiale}$) par le nombre de colonnes ($NbreColonnes = \frac{LargeurChampCellulaire}{ResolutionSpatiale}$) de la grille du carroyage appliqué à la zone d'étude. Chaque cellule pourra ainsi être identifiée (C_{ij} avec $0 \leq i \leq NbreLignes - 1$ et $0 \leq j \leq NbreColonnes - 1$) par sa position topologique dans la grille à partir de la ligne et de la colonne dont le croisement définit son coin supérieur gauche comme l'illustre la figure 6.4.

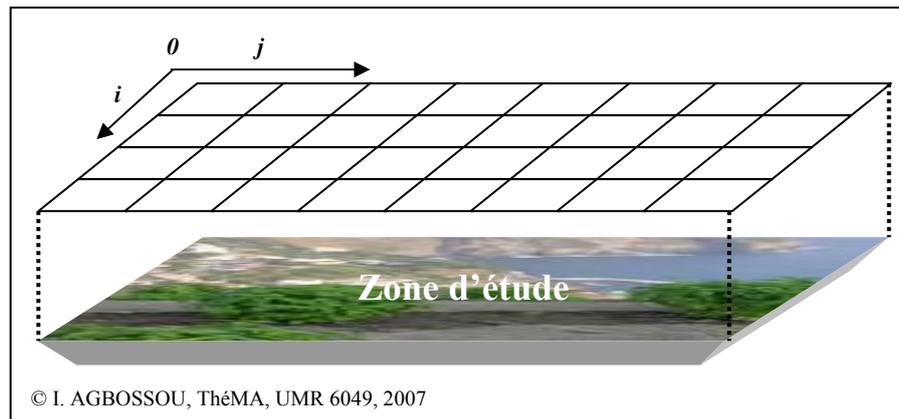


Figure 6.4 : Application du carroyage à la zone d'étude

- *Les états* : chaque état correspond à une occupation du sol. Il s'agit de : appartement loué, appartement acheté, maison louée, maison achetée, eau, commerce, équipement, réseau routier ou ferré, espace vert ou forêt, zone constructible (que nous notons respectivement *AptLoue*, *AptAchete*, *MLouee*, *MAcheee*, *Eau*, *Cce*, *Eq*, *Reseau*, *EspaceVert*, *ZoneConst*). Nous distinguerons trois catégories d'états : les états dynamiques, les états statiques et les états pseudo-dynamiques. Les états dynamiques matérialisent les occupations du sol susceptibles de changer de nature des le temps c'est-à-dire pendant la simulation (*AptLoue*, *AptAchete*, *MLouee* et *MAcheee*). Les états statiques (*Eau*, *Cce*, *Eq*, *Reseau* et *EspaceVert*) ne subissent pas de transition au fil du temps mais exerce une certaine influence sur le comportement des ménages en ce qui concerne leurs choix résidentiels. L'état pseudo-dynamique (*ZoneConst*) se comporte comme un état statique tant qu'il n'y a pas de nouvelle dans la cellule qu'il occupe ; mais se comporte comme un état dynamique dans le cas contraire. L'ensemble des états du système est donc l'ensemble E défini par $E = \{AptLoue, AptAchete, MLouee, MAcheee, Eau, Cce, Eq, Reseau, EspaceVert, ZoneConst\}$
- *Le voisinage* : comme justifié au chapitre 5, c'est au travers du voisinage de Moore (chaque cellule possède théoriquement huit voisins) que nous matérialisons par la concrétisation de l'interaction spatiale entre les cellules. En réalité, toutes les cellules n'auront pas exactement huit voisins en raison de leur position topologique dans le champ cellulaire. C'est le cas par exemple des cellules C_{00} , C_{0j} , C_{i0} et $C_{NbreLignes-1, NbreColonnes-1}$. L'interaction spatiale des cellules est donc modélisée, pour chaque cellule, par un potentiel d'attraction ($P_{C_{ij}}$)

mesuré par le rapport entre le nombre de cellules de son voisinage ayant le même état qu'elle-même, pondéré par un facteur d'accélération ou de freinage et le nombre de cellules présentes dans le voisinage.

$$P_{C_{ij}} = \frac{\sum VoisinsDeMemeEtat(C_{ij})}{NbreVoisins} * \alpha_{ij} \quad (6.5)$$

La fonction $\sum VoisinsDeMemeEtat(C_{ij})$ retourne par calcul le nombre de cellules ayant le même état que la cellule C_{ij} dans son voisinage. Le facteur d'accélération ou de freinage α_{ij} est obtenu par l'application d'une fonction sigmoïde :

$$\frac{1}{1 + e^{-\lambda t}}$$

où t désigne le temps et le paramètre λ la qualité de l'environnement résidentiel. En effet, les propriétés de la fonction sigmoïde s'expliquent par celle de sa dérivée

$$\frac{d\alpha_{ij}}{dt} = \frac{\lambda * e^{-\lambda t}}{(1 + e^{-\lambda t})^2}$$

qui peut se transformer en

$$\frac{d\alpha_{ij}}{dt} = \lambda * (\alpha_{ij}) * (1 - \alpha_{ij}).$$

Cette équation différentielle exprime le fait que la variation de α_{ij} dans l'intervalle $]0, 1[$ au cours du temps est à la fois proportionnelle à elle-même (α_{ij}) et à son complément ($1-\alpha_{ij}$), proportionnalité affectée par la qualité environnementale. La figure 6.5 (où $f(x) \equiv \alpha_{ij}$ et $x \equiv -\lambda t$) donne la forme de la fonction sigmoïde.

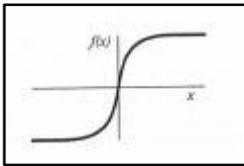


Figure 6.5 : Allure d'une fonction Sigmoïde

- *La configuration initiale* : le point de départ de toute simulation est une configuration initiale définissant l'occupation du sol de chaque cellule. Dans notre cas, elle peut être fournie par l'utilisateur du simulateur grâce à un éditeur d'état embarqué dans le simulateur ou encore être définie de façon aléatoire à

partir de la cartographie de la zone d'étude. Le comportement résidentiel des agents (ménages) modifie à la première itération cette configuration pour en faire une autre. Et c'est le croisement de ces deux configurations diachroniques (la configuration initiale et celle générée par le comportement des ménages à la première itération) qui permet d'établir la table de contingence, laquelle permet de construire la matrice de transition. Cette construction diachronique de la matrice de transition se fait donc de façon dynamique durant toute la durée de la simulation entre les dates $t-2$ et $t-1$ pour à l'occupation du sol à la date t .

6.3.2. Construction de la matrice de transition

Une table de contingence permet de représenter de façon statique l'information dans un tableau à double entrée. La prise en compte de la dimension dynamique passe par la transformation de cette table en matrice de transition puis en chaîne de Markov.

Pour acquérir les informations nécessaires à l'élaboration de la table de contingence, nous proposons un algorithme (AGTCS : *Algorithme de Génération de Table de Contingence Spatiale*) qui prend comme données d'entrée la configuration initiale. La zone d'étude est partitionnée en autant de classes qu'il y a d'états, en l'occurrence nous aurons dix classes (c'est-à-dire dix partitions spatiales de la zone d'études). Chaque classe est alors étiquetée par un état. On comptabilise ensuite le nombre d'occurrences (chaque cellule matérialisant un état donné) présentes pour chaque état dans chacune des classes. La table de contingence ainsi obtenue possède autant de lignes et de colonnes qu'il y a d'états dans le système. Dans notre cas cette table possède dix lignes et dix colonnes comme schématisé à la figure 6.6.

| | Classes de partition | | | |
|----------|----------------------|------------|-------|-----------|
| États | I | II | | X |
| e_1 | n_{1I} | n_{1II} | | n_{1X} |
| e_2 | n_{2I} | n_{2II} | | n_{2X} |
| ... | ... | ... | | ... |
| e_{10} | n_{10I} | n_{10II} | | n_{10X} |

Figure 6.6 : Structure de la table de contingence

Les informations contenues dans la partie grisée de la table représentent le nombre de cellules se trouvant dans un état donné dans une partition précise. Ainsi la somme de tous les éléments de chaque ligne ($n_i = \sum_{j=1}^X n_{ij}$) correspond au nombre total de cellules occupant l'état e_i dans la configuration initiale. La somme des n_i donne le nombre total de cellules définissant le système. L'objectif étant d'obtenir une matrice carrée contenant les valeurs constituant chaque vecteur d'état, nous appliquons le principe de la fonction *Max* sur chaque colonne afin de l'identifier par l'étiquette de l'état dominant. La figure 6.7 déroule l'algorithme en pseudo code permettant de calculer la table de contingence à partir de la configuration initiale.

```

fonction AGTCS(ConfigurationInitiale) retourne la table de contingence
  Données d'entrée :
    ConfigurationInitiale ; // définie à partir de la zone d'étude par l'utilisateur
  Variables locales :
    NbreEtats ← card(E) ; // E étant l'ensemble des états du système
    NbreDeCellulesParPartition ← NbreTotalCellules/NbreEtats ;
    Partition ← Liste<Cellule> ; // création d'une structure de données de
    // liste de cellules pour chaque partition
    Parties ← Liste<Partition> ; // création d'une liste de NbreEtats éléments
    // pour contenir les partitions
    Etats ← Liste<Etat> ; // création d'une liste de NbreEtats éléments
    // pour stocker l'ensemble des états
    Table ← double[NbreEtats][NbreEtats] ; // création d'un tableau à
    // double entrée pour stocker
    // les informations

  Calcul des effectifs  $n_{ij}$  :
    Pour chaque Partition  $p_j$  de Parties
      Pour chaque Etat  $e_i$  de Etats
        Partition ←  $p_j$  ;
        occurrence ← 0 ;
        Pour chaque cellule  $c$  de  $p_j$ 
          Si  $e_i ==$  état de  $c$ 
            occurrence++ ;
          FinSi
        FinPour
        Table[i][j] ← occurrence ;
      FinPour
    FinPour
  Retourne Table ; // retourne la table de contingence remplie
  
```

© I. AGBOSSOU ThéMA, UMR 6049, 2007

Figure 6.7 : Algorithme de Génération de Table de Contingence Spatiale

Dans cet algorithme, nous devons signaler que le partitionnement de la zone d'étude en classes (étiquetées chacune par un état) est basé sur la méthode des stratifications

notamment le sondage en grappes à probabilités égales. En effet, la stratification consiste en des tirages séparés effectués dans des sous-populations supposées, en l'occurrence, plus homogènes que la population mère (la zone d'étude elle-même). Etant donné que nous travaillons sur un espace géographique, l'ensemble des strates constitue des grappes [Saporta, 2006 ; Robert, 1996 ; Droesbeke *et al.*, 2002] de dix sous-populations.

Une fois la table de contingence obtenue il devient plus aisé de calculer la matrice de transition par transformation des effectifs n_{ij} en une valeur probabiliste $f_{ij} = n_{ij}/n_i$.

```

fonction MatriceDeTransition(tableDeContingence) retourne la matrice de transition
Données d'entrée :
    tableDeContingence ; // fournie par la méthode AGTCS
Variables locales :
    NbreEtats ← card(E) ; // E étant l'ensemble des états du système
    ConfigurationInitiale ; // définie à partir de la zone d'étude
    ni ; // nombre de cellules de chaque état (i = 1, ..., 10) présent dans
        // dans la configuration initiale du système cellulaire
    Table ← double[NbreEtats][NbreEtats] ; // création d'un tableau à double
        // entrée pour stocker les valeur fij

Calcul des valeurs probabilistes fij :
    Pour j = 1 à 10
        Pour i = 1 à 10
            fij ← TableDeTransitio[i][j] / ni ;
            Table[i][j] ← fij ;
        FinPour
    FinPour
Retourne Table ; // retourne la de matrice de transition remplie

```

La matrice de transition se présente sous la forme suivante :

$$Q = \begin{bmatrix} f_{11} & \cdots & f_{1,10} \\ \vdots & \ddots & \vdots \\ f_{10,1} & \cdots & f_{10,10} \end{bmatrix}$$

© I. AGBOSSOU Théma, UMR 6049, 2007

Figure 6.8 : Calcul de la matrice de transition

L'estimation d'une matrice de transition à partir d'un ensemble de données amène à se poser au moins une question à savoir "La matrice a-t-elle été correctement estimée ?". Si le calcul des valeurs probabilistes de la matrice ne pose pas de problème particulier, rien ne permet *a priori* d'affirmer que l'estimation obtenue est de bonne qualité. Concrètement, il est nécessaire de se demander si les données utilisées permettent une estimation robuste de la matrice. En effet, une matrice de transition est un objet synthétisant l'information sur les relations entre plusieurs variables. Pour quantifier

cette information, il est nécessaire de définir un principe de mesure. En l'occurrence, nous prendrons pour point de départ l'idée qu'une matrice est d'autant plus informative, dans le cadre de l'application des chaînes de Markov, que la connaissance de la modalité ligne permet de réduire l'incertitude existant sur la modalité colonne [Meyn et Tweedie, 1993 ; Berchtold, 1998, *op. cit.* ; Dupuis, 2002]. Aussi l'une des approches les plus pertinentes dans ce cas est l'utilisation d'une matrice de transition d'ordre 1 calculée à partir de la représentation autorégressive [Raftery et Tavaré, 1994] des données (la configuration initiale dans notre cas) ; voilà qui justifie la pertinence de l'algorithme de génération de la table de contingence spatiale.

La dynamique visuelle du système est régie par les règles de transition des cellules (passage d'un état à un autre). Mais elle est également soumise aux variations comportementales de l'ensemble des ménages qui en retour modifient les configurations spatio-fonctionnelles de l'habitat par le biais de jeux d'interactions spatio-temporelles. La modélisation du couplage des SMA et des automates cellulaires nous permet de préciser d'avantage le déroulement de l'ensemble de ces processus.

6.4. Modèle algorithmique du système résidentiel : couplage SMA-AC

Si l'on avait une connaissance parfaite de l'environnement dans lequel nous vivons, l'approche logique devrait permettre de faire des choix ou de confectionner des plans dont le bon fonctionnement serait garanti. Malheureusement, les agents n'ont pratiquement jamais accès à toute la connaissance sur leur environnement, et doivent donc agir dans l'incertitude.

6.4.1. Comment modéliser le comportement humain ?

Si l'on s'intéresse au comportement humain, il est nécessaire de se pencher sur un certain nombre de sujets tels que la compréhension des mécanismes sous-tendant, entre autres, le fonctionnement de la mémoire et de la perception. En somme, arriver à comprendre et décrypter le fonctionnement des facultés cognitives. D'après H. Mallot [Mallot, 1997], le mot cognition se réfère généralement à des processus mentaux variés tels que l'attention, la reconnaissance des lieux et des objets, la planification et le raisonnement. La théorie du contrôle en psychologie comportementale, telle que la

décrivent R. Lord et P. Levy [Lord et Levy, 1994] reprend le principe des boucles de rétroaction tout en l'étendant à l'ensemble des processus comportementaux de la tâche d'asservissement à la régulation des comportements sociaux. Pour ces chercheurs, la généralité des boucles de rétroaction pour la description du comportement provient de la nature hiérarchique des systèmes de contrôle, même si la nature des activités de contrôle peut être très différente selon les niveaux. Parmi les diverses classifications de ces niveaux, nous nous référons à celle introduite par A. Newell dans *Unified Theories of Cognition* [Newell, 1990]. Cette classification est décrite en quatre niveaux : le premier représente des comportements réflexes de type attraction-répulsion, le second représente des comportements nécessitant une intégration spatio-temporelle, le troisième traite de la plasticité du second et le dernier, appelé cognitif et qui nous intéresse en l'occurrence, ne dépend plus seulement de stimuli sensoriels, mais aussi des buts poursuivis par la personne.

La problématique de la modélisation du comportement humain dans le cadre d'applications de simulation n'est pas de reproduire fidèlement l'intelligence humaine mais de proposer une architecture permettant de modéliser des comportements crédibles et réalistes d'agents anthropomorphes évoluant en temps réel, chacun, dans son environnement.

6.4.2. Revisite de l'architecture d'agent cognitif de type BDI

Les systèmes réactifs ne sont pas suffisants pour décrire l'ensemble des comportements, notamment rationnels et sociaux. Il est ainsi nécessaire de recourir aux modèles cognitifs qui représentent des connaissances (croyances, intentions, plans objectifs, etc.) d'un agent. La figure 4.2 du quatrième chapitre, sert de support pour représenter l'architecture d'agents cognitifs. Les architectures cognitives visent à unifier au sein d'un même modèle un ensemble de propriétés communément admises au sein de la communauté scientifique (psychologie, neurosciences, ergonomie, intelligence artificielle et intelligence artificielle distribuée, animation et réalité virtuelle, etc.) sur la cognition humaine [Lamarche, 2003 ; Donikian, 2004 ; Burkhardt et al., 2004]. Aussi, le comportement d'un agent dépend de son espace d'évolution et de ces connaissances. En adéquation avec la théorie de Gibson [Gibson, 1986] et plus précisément celle de Rillieux et Quéré [Rillieux et Quéré, 1998] sur les affordances spatiales, G. Thomas [Thomas, 2005] a montré que l'on peut informer l'environnement urbain en

caractérisant les espaces d'évolution, en prenant en compte leur configuration, les objets qu'ils contiennent et leur structuration méreotopologique (c'est-à-dire la combinaison de la topologie à la méréologie permettant ainsi d'ajouter à la relation de l'ensemble du voisinage d'une cellule donnée à la zone d'étude entière). L'environnement contient donc les informations nécessaires aux prises de décision des agents. Ceci est illustré par l'algorithme générique de prise de décision atomique (une action) d'un agent dans la figure 6.9. Conformément à l'architecture BDI, les intentions permettent à un agent de raisonner sur son état interne et celui des autres avec lesquels il est en interaction à partir des informations qu'il obtient par le biais de son environnement. Avec ce modèle, il est possible de spécifier les buts à atteindre et de trouver une stratégie les satisfaisant. La simulation par agents autonomes dotés de comportements sophistiqués passe donc par la prise en compte de l'environnement. En complément de la représentation géographique et géométrique (modélisée par l'automate cellulaire) de l'environnement, il est nécessaire de fournir pour chaque agent un modèle symbolique des autres agents avec lesquels il entretient une relation sociale.

```

fonction DecisionAtomique(environnement) retourne une action
  Données d'entrée :
    environnement ; //environnement dans lequel se trouve l'agent
  Variables locales :
    Agents ; // les autres agents avec lesquels il est
              // en interaction
    Desirs ; // les désirs possibles de l'agent
    But ; // un but, initialement vide
    Action ; // action à accomplir, initialement vide

    Agents ← GetAlterEgo(environnement) ;
    Desirs ← ActualiserMesDesirs(environnement, Agents) ;
    But ← FormulerMonBut(Desirs) ;
    Action ← IdentifierLesActions(But) ;
  Retourne Action ;

```

© I. AGBOSSOU ThéMA, UMR 6049, 2007

Figure 6.9 : Algorithme de prise de décision atomique d'un agent BDI résidentiel

Chaque agent, à chaque itération, commence par actualiser ses désirs ; ce qui lui permet de formuler ou de reformuler un nouvel objectif à atteindre. Pour ce faire, il identifie la ou les actions à entreprendre. Pour un ménage, une action peut prendre l'une des valeurs suivantes : chercher un appartement à louer, chercher un appartement à acheter, chercher une maison à louer ou chercher une maison à acheter.

Mais le fait d'identifier des actions à accomplir n'induit pas automatiquement une prise de décision de la part d'un agent résidentiel quand à son changement de logement. En effet, d'autres contraintes et conditions sont requises pour parvenir à une prise de décision en ce qui concerne le choix et la migration.

6.4.3. Le choix d'un nouveau lieu de résidence et la décision de migrer

Dès lors que le ménage a identifié l'action à accomplir, le processus de mobilité peut commencer. Trois éléments particulièrement intéressants entrent en jeu : la propension du ménage à être mobile, le choix du nouveau lieu de résidence et la migration effective. Détaillons pas à pas la formalisation de chacun de ces trois éléments.

- **La propension à migrer** : elle est une fonction globale de l'ensemble des variables caractéristique d'un ménage. Ainsi, à partir de la base de connaissance (cf. Figure 6.3) d'un agent, il est possible de calculer sa propension à migrer en déroulant l'inférence bayésienne suivante : $P_k^t = P(\text{migration}|K)$ où k désigne un type de ménage donné et K le vecteur de modalités décrivant le ménage à la date t .
- **Le choix du nouveau lieu de résidence** : en ce qui concerne le choix de la localisation résidentielle, le comportement migratoire des ménages est fortement influencé par les caractéristiques (la qualité environnementale) et les propriétés intrinsèques des logements actuel et futur. Ainsi, pour orienter son choix, le ménage a besoin d'espace de référence muni d'une certaine métrique ; c'est-à-dire les limites spatiales de son champ d'investigation. Il s'agit en l'occurrence des limites spatiales définies par le concept de voisinage que nous avons appliqué aux cellules. En effet, l'agent procède à une exploration systématique de la zone d'étude selon un partitionnement basé sur le voisinage. En fonction de l'offre de logements disponibles, il sélectionne les logements susceptibles de l'accueillir en ne retenant que ceux dont le potentiel d'attraction ($P_{C_{ij}}$) est supérieur à son confort résidentiel actuel ($P_k^t * \lambda$). Notons que par souci de simplification et de réalisme, le paramètre λ qui exprime la qualité de l'environnement résidentiel (cf. §5.3.1) est précisé en entrée au modèle de simulation par l'utilisateur pour chaque catégorie de logement. Enfin, une comparaison deux à deux de ces logements lui permet de choisir celui qui a le

meilleur score. C'est ce logement qui est finalement susceptible de l'accueillir, s'il migre effectivement.

- **La migration effective** : il n'est pas rare que l'on revienne sur une décision ou un choix pour une quelconque raison. Aussi dans le cadre du modèle de simulation que nous élaborons, étant donné que la résidence seconde n'est pas prise en compte, un changement de logement n'est effectif que si le nouveau logement choisi est occupé par le ménage et que l'ancien est libéré.

Pour une vision synthétique de l'ensemble des processus régissant à la fois le comportement résidentiel des ménages et leurs influences réciproques sur l'habitat dont la dynamique est modélisée selon une approche par automates cellulaire. Nous présentons le modèle mathématique de l'évolution du système résidentiel.

6.4.4. Formalisation mathématique de la dynamique globale du système

Avant de pouvoir présenter le modèle global du système, nous donnons quelques précisions sur sa dynamique démographique. Pour que l'évolution de la structure des ménages simulée dans le système prolonge de manière cohérente celle observée par l'Insee depuis une vingtaine d'années, il convient de modéliser de façon relativement précise l'histoire familiale des ménages. Il s'agit en l'occurrence de la formation d'un nouveau ménage, l'éclatement d'un ménage existant et la disparition d'un ménage. Pour ce faire nous nous basons sur la typologie des ménages à quatre niveaux (personne seule, couple sans enfant, couple avec enfant et famille monoparentale) que nous avons présentée au deuxième chapitre (*cf.* figure 2.6).

La modélisation des événements est souvent délicate car les données disponibles restent parcellaires et permettent difficilement de rendre compte des effets de génération. Pour cela et en raison d'autres facteurs d'instabilité à long terme des modèles, les durées de simulation n'excéderont pas la décennie. Les choix méthodologiques retenus pour la modélisation de la dynamique démographique du système ont été de partir des résultats de microsimulation dynamique du modèle DESTINIE [INSEE, 1999 *op. cit.* ; Robert-Bobée, 2002a *op. cit.* ; Bonnet et *al.*, 1999] élaboré par des services de L'Insee et aussi des tables de mortalité françaises pour les XIX^{ème} et XX^{ème} siècles et projections pour le XXI^{ème} siècle [Vallin et Meslé, 2002]. Les comportements matrimoniaux et la fécondité intégrés dans DESTINIE sont estimés à partir de l'enquête Jeunes et Carrières, enquête complémentaire à l'enquête Emploi réalisée en mars 1997 par l'Insee. Ceci a conduit à

la reproduction, de probabilités, entre autres, de mise en couple, de ruptures des unions [Robert-Bobée, 2001 ; 2002b].

6.4.4.1. La formation de nouveaux ménages : union ou éclatement d'un ménage existant

Les statistiques d'état civil offrent un point de départ à la modélisation. Elles fournissent des taux de nuptialité qui sont ensuite "adaptés" pour tenir compte du comportement des ménages. On dispose notamment des taux conjoncturels de primo-nuptialité en fonction de l'âge et du sexe [Kerjosse et Tamby, 1996 ; Toulemon, 1996]. Selon ces sources, la première mise en couple se situerait entre 16 et 49 ans pour les femmes avec une valeur modale de 26 ans, entre 18 et 50 ans pour les hommes avec un pic à 28 ans. Ces statistiques d'état civil ne peuvent toutefois rendre compte à elles seules des comportements matrimoniaux des ménages. En effet, la vie maritale s'est largement diffusée comme mode d'entrée en union depuis quatre décennies ans, passant de 10 % à 90 % des couples entre 1965 et 1995 [Villeneuve-Gokalp, 1994]. Si la cohabitation hors mariage est la forme de vie en couple principale en début d'union, particulièrement chez les jeunes, elle devient aussi un mode de vie durable [Lefranc, 1995 ; Daguet, 1996]. Aussi, dans le cadre de ce travail, nous traitons indifféremment les mises en couple suite au mariage, les cohabitations hors mariage ainsi que les couples homosexuels car chaque ménage d'un type donné est représenté par une personne de référence. La modélisation retenue à ce stade consiste à prendre, à chaque pas de temps, un taux $T_{\text{UnionDesPersonnesSeules}}$ (dont la valeur par défaut est de 90%, mais paramétrable en début de simulation) des ménages de type "Personne seule", les mettre en union deux à deux pour former de nouveaux ménages de type "Couple sans enfant". Le départ du domicile parental peut être ou non consécutif à une mise en couple. En effet, nous faisons l'hypothèse qu'il y a éclatement d'un ménage existant seulement lorsqu'un enfant en début de vie active quitte ses parents. Les ménages dont la personne de référence ou le conjoint éventuel est étudiant ou en début de vie active représentent moins de 2 % de l'ensemble des ménages ordinaires (Insee, enquête « Budgets de Famille » 1995). Ainsi nous appliquons un taux $T_{\text{Décohabitation}}$ (dont la valeur par défaut est de 2%, mais paramétrable en début de simulation) sur l'ensemble des ménages de types "Famille monoparentale" et "Couple avec enfants" pour obtenir de nouveaux ménages de type "Personnes seules" à chaque pas de temps durant la simulation. Ce

choix de modélisation présente l'avantage que les jeunes ménages présentent des revenus nuls (car on ne modélise pas la solidarité financière entre ménages).

Outre le départ des enfants du domicile parentale, il y aussi les dissolutions des couples. Les statistiques d'état civil fournissent des taux de divorce ou de séparation en fonction de la durée du mariage ou de la cohabitation. Selon ces sources, l'intensité finale des séparations est égale à 35%, répartie entre 1 et 40 ans d'union, avec un mode estimé à 3 ans [Sardon, 1996 ; Daguet, 1996, *op. cit.*]. Compte tenu du faible taux de divorce observé au delà de 40 années de mariage, nous supposons qu'un couple ne se sépare plus au delà de quarante ans de vie commune. Mais dans le cadre de ce travail, étant donné que la durée maximale de simulation est limitée à dix ans avec un pas de temps calé sur un semestre, aucun ménage n'aura accumulé quarante de vie commune à la fin de toute simulation. Et nous appliquons un taux de dissolution $T_{Dissolution}$ (dont la valeur par défaut est de 35%, mais paramétrable en début de simulation) aux ménages de types "Couple sans enfant" et "Couple avec enfants". Ainsi, si un couple sans enfant se dissout, alors l'effectif des ménages de type "Personne seule" est incrémenté de deux point et celui des "Couple sans enfant" est décrémenté de l'unité. En revanche, s'il s'agit d'un couple avec enfant, alors la catégorie "Personne seule" est incrémentée d'un point, la catégorie "Famille monoparentale" d'un point aussi et "Couple avec enfant" est décrémenté de l'unité. Les équations algorithmiques de la figure 6.6 résument l'ensemble de ces dynamiques.

```

EffeciftPersonnesSeules -= TUnionDesPersonnesSeules * EffeciftPersonnesSeules ;
EffeciftPersonnesSeules += (EffeciftCoupleAvecEnfants + EffeciftFamilleMonoparentale) * TDécohabitation ;
EffeciftCoupleSansEnfant +=  $\frac{1}{2}$ TUnionDesPersonnesSeules * EffeciftPersonnesSeules ;

// Dissolution des couples sans enfant
EffeciftCoupleSansEnfant -= TDissolution * EffeciftCoupleSansEnfant ;
EffeciftPersonnesSeules += 2 * TDissolution * EffeciftCoupleSansEnfant ;

// Dissolution des couples avec enfants
EffeciftCoupleAvecEnfant -= TDissolution * EffeciftCoupleAvecEnfant ;
EffeciftPersonnesSeules += TDissolution * EffeciftCoupleAvecEnfant ;
EffeciftFamilleMonoParentale += TDissolution * EffeciftCoupleAvecEnfant ;

```

© I. AGBOSSOU ThéMA, UMR 6049, 2007

Figure 6.10 : Equations algorithmiques de la formation de nouveaux ménages

6.4.4.2. La disparition d'un ménage

La disparition d'un ménage s'exprime par le fait qu'il n'a plus d'instance dans le système. Cela veut dire que la personne de référence est décédée. Cette modélisation peut s'effectuer de deux façons. D'une part, on peut simplement considérer l'espérance de vie des personnes de référence des ménages. D'autre part, on peut estimer le quotient de mortalité *ex-ante* des individus et l'appliqué à la personne de référence. Dans ce cas, ce quotient de mortalité peut faire office d'un indicateur de morbidité explicatif des probabilités de décès. Cependant, pour faire simple, nous choisissons dans le cadre de notre travail, la première option. Ainsi le décès de la personne de référence n'aura d'effet que sur les ménages de type "Personne seule". Et le logement occupé par ce dernier est donc automatiquement libéré. Les autres types de ménages ne subiront aucune influence de la disparition de leur personne de référence dans le sens où nous faisons l'hypothèse qu'il y aura soit substitution de cette dernière en cas de décès soit absence de désir de déménager. Dans ce cas, seule l'équation algorithmique suivante est requise pour la simulation :

$$\begin{cases} \text{Pour chaque ménage de type "Personne seule"} \\ \text{Si } Age == EsperanceDeVie, \text{ alors} \\ Effectif_{PersonneSeule}^t \leftarrow Effectif_{PersonneSeule}^t - 1; \end{cases} \quad (6.6)$$

Comme pour les divers taux démographiques, l'espérance de vie des ménages est aussi paramétrable en début de simulation.

6.4.4.3. Mathématisation de la dynamique du système résidentiel

Eu égard à tout ce qui précède, nous représentons dans ce paragraphe le formalisme mathématique régissant l'ensemble du système résidentiel. En notant M_k un ménage de type $k \in \{\text{"Personne seule"}, \text{"Couple sans enfant"}, \text{"Couple avec enfant"}, \text{"Famille monoparentale"}\}$ et $C_{ij}(e)$ une cellule donnée dans un état $e \in \{\text{AptLoue}, \text{AptAchete}, \text{MLouee}, \text{MAchetee}, \text{Eau}, \text{Cce}, \text{Eqt}, \text{Reseau}, \text{EspaceVert}, \text{ZoneConst}\}$, on peut formaliser l'état du système résidentiel à une date t de la façon suivante :

$$\begin{cases} M_k^t(l) = f \left(M_k^{t-\delta t}(l), P_k^{t-\delta t}, \bigcup_{i=0; j=0}^{n; m} C_{ij}^{t-\delta t}(e) \right) \\ C_{ij}^t(e) = g \left(C_{ij}^{t-\delta t}(e), P_{C_{ij}}^{t-\delta t}, M_k^t(l) \right) \end{cases} \quad (6.7)$$

δt désigne le pas de temps (par défaut chaque itération correspondra à une durée de six mois dans la réalité) ;

$\bigcup_{i=0; j=0}^{n; m} C_{ij}^{t-\delta t}(e)$ modélise l'offre de logements disponibles et c'est parmi ces derniers que le ménage de type k pourra choisir un logement l à la date t si sa migration est effective : $M_k^t(l)$. Dans ce cas son ancien logement devra se libérer et s'ajouter à l'offre disponible ;

$C_{ij}^t(e)$ formalise la transition d'une cellule donnée d'un état à un autre, et ce en fonction de l'état dans lequel elle se trouve à l'itération précédente, de son potentiel d'attraction et du choix résidentiel des ménages.

En effet, les équations (6.7) traduisent la double interaction entre les deux grandes entités du système résidentiel (*cf.* Figure 2.2) à savoir l'habitat et les habitants. A chaque pas de temps δt , il se produit une série d'interactions (changement de statut d'occupation, achat, location de logement, vieillissement, etc.) au niveau des ménages (SMA). Ces interactions ont une répercussion spatiale qui se matérialise au niveau de l'automate cellulaire. Celui-ci, en retour, rétroagit sur les comportements des ménages. Le processus évolue donc de cette façon durant toute la durée de simulation. La transformation de l'équation (6.7) conduit à l'obtention de l'ensemble des règles probabilistes de transition qui devraient gouverner la dynamique résidentielle du système.

En posant l'hypothèse markovienne selon laquelle "*la probabilité de transition d'une cellule d'un état à un autre ne dépend que de son état actuel et non de son histoire*"

$$\begin{cases} P_k^{t+1}(C_{ij}^{t+1} = e) = P_k(C_{ij}^{t+1} = e | C_{ij}^t = e') * P(C_{ij}^t = e') \\ P(C_{ij}^{t+1} = e) = \sum_{ij}^{voisinage} P(C_{ij}^t = e | C_{ij}^{t-1} = e' \in E \setminus \{e\}) \end{cases} \quad (6.8)$$

La première ligne de l'équation (6.8) se traduit comme suit : la probabilité pour qu'un ménage de type k choisisse un logement dont l'état est e au temps $t+1$. La seconde ligne formalise la probabilité qu'une cellule occupe l'état e au temps $t+1$ d'après son voisinage où s'exprime son potentiel d'attraction. La figure 6.12 et la figure 6.11 présentent respectivement le code source de chacune de ces lignes d'équation. Dans la

figure 6.12 on remarquera l’usage d’un générateur pseudo-aléatoire (la fonction *Random*) sur les cellules. En effet, lorsque plusieurs cellules remplissent les conditions de choix pour un ménage donné, ce dernier choisit une au hasard.

```
// Règles de transition intrinsèques des cellules selon le voisinage et le potentiel
foreach(Cellule cel in CellulesSimulables)
{
    // récupérer le potentiel d'attraction de la cellule au temps t+dt
    double potentiel = cel.PotentialOfAttraction(CompteurDuTemps);
    //Transition d'une cellule constructible
    if(cel.IsConstructible)
        switch(cel.GetEtatDominantDuVoisinage())
        {
            case EtatPossible.AppartementAchete:
                cel.ChangeToAppartementAchete(CleValeurEtatCouleur);
                break;
            case EtatPossible.AppartementLoue:
                cel.ChangeToAppartementLoue(CleValeurEtatCouleur);
                break;
            case EtatPossible.MaisonAchete:
                cel.ChangeToMaisonAchete(CleValeurEtatCouleur);
                break;
            case EtatPossible.MaisonLoue:
                cel.ChangeToMaisonLoue(CleValeurEtatCouleur);
                break;
        }
    else if(cel.Etat == EtatPossible.MaisonLoue && potentiel > 1.0)
    {
        // Si l'état dominant du voisinage est une maison achetée alors, changer en maison achetée
        if(cel.GetEtatDominantDuVoisinage() == EtatPossible.MaisonAchete)
            cel.ChangeToMaisonAchete(CleValeurEtatCouleur);
    }
    else if(cel.Etat == EtatPossible.MaisonAchete && potentiel > 1.0)
    {
        // Si l'état dominant du voisinage est une maison louée alors, changer en maison louée
        if(cel.GetEtatDominantDuVoisinage() == EtatPossible.MaisonLoue)
            cel.ChangeToMaisonLoue(CleValeurEtatCouleur);
    }
    else if(cel.Etat == EtatPossible.AppartementLoue && potentiel > 1.0)
    {
        // si l'état dominant du voisinage est appartement acheté alors, change en appartement acheté
        if(cel.GetEtatDominantDuVoisinage() == EtatPossible.AppartementAchete)
            cel.ChangeToAppartementAchete(CleValeurEtatCouleur);
    }
    else if(cel.Etat == EtatPossible.AppartementAchete && potentiel > 1.0)
    {
        // Si l'état dominant du voisinage est appartement loué alors, change en appartement loué
        if(cel.GetEtatDominantDuVoisinage() == EtatPossible.AppartementLoue)
            cel.ChangeToAppartementLoue(CleValeurEtatCouleur);
    }
}
}
```

Figure 6.11 : *Algorithme des règles de transition selon le voisinage et le potentiel*

```

// Règles de transition dictées par le comportement des ménages

//Récupération des différents typs de logements (maisons ou appartements)
List<Cellule> maisons = EnsembleDesLogements.FindAll(Cellule.EstUneMaison);
List<Cellule> apparts = EnsembleDesLogements.FindAll(Cellule.EstUnAppartement);

foreach(Menage agent in AgentsResidentiels.EnsembleDesMenages)
{
    switch(agent.DesirDuMenage)
    {
        case Desir.HabiterUnAppartement:
        {
            Cellule cellule = apparts[new Random().Next(0, apparts.Count - 1)];
            if(agent.ActionsPossibles.Contains(Action.AppartementAchete)
                && agent.Statut == StatutOccupation.Proprietaire)
            {
                cellule.ChangeToAppartementAchete(CleValeurEtatCouleur);
                agent.MigrerEffectivementVers(cellule);
            }
            else if(agent.ActionsPossibles.Contains(Action.AppartementLoue)
                && agent.Statut == StatutOccupation.Locataire)
            {
                cellule.ChangeToAppartementLoue(CleValeurEtatCouleur);
                agent.MigrerEffectivementVers(cellule);
            }
        }
        break;
        case Desir.HabiterUneMaison:
        {
            Cellule cell = maisons[new Random().Next(0, maisons.Count - 1)];
            if(agent.ActionsPossibles.Contains(Action.MaisonAchete)
                && agent.Statut == StatutOccupation.Proprietaire)
            {
                cell.ChangeToMaisonAchete(CleValeurEtatCouleur);
                agent.MigrerEffectivementVers(cell);
            }
            else if(agent.ActionsPossibles.Contains(Action.MaisonLouee)
                && agent.Statut == StatutOccupation.Locataire)
            {
                cell.ChangeToMaisonLouee(CleValeurEtatCouleur);
                agent.MigrerEffectivementVers(cell);
            }
        }
        break;
    }
}

```

Figure 6.12 : *Algorithme des règles de transition selon le comportement des ménages*

L'objectif de ce chapitre a été de montrer et d'expliquer les différentes étapes (les plus importantes à notre avis) d'élaboration de notre modèle de simulation (*VisualSimores* dont les fonctionnalités logicielles sont détaillées au chapitre suivant) selon une approche mathématique et informatique. Après avoir précisé les bases fondamentales du raisonnement par calculs de probabilités conditionnelles sur plusieurs variables simultanées, nous avons défini les conditions d'application de ce principe de l'inférence bayésienne dans la démarche de formalisation des comportements des agents cognitifs (les ménages) en ce qui concerne le processus du choix résidentiel. Ceci nous a permis d'aboutir à l'équation dynamique de la probabilité de changer de logement pour chaque catégorie de ménage. Par ailleurs les aspects spatiaux ont été également pris en compte au travers de la spécification d'un automate cellulaire dont les règles de transition sont contraintes par le comportement cognitif des ménages. Finalement nous avons abouti à la formalisation du couplage SMA-AC qui exprime les dynamiques spatiales et temporelles du système résidentiel.

Ce couplage SMA-AC est intégré et programmé sous forme de prototype logiciel pour des fins de tests et de simulations pour générer des scénarios dont les résultats sont présentés au septième et dernier chapitre de notre thèse.

Bibliographie référencée

- [**Bonnet et al., 1999**] C. Bonnet, C. Burricaud, C. Colin, A. Flipo, R. Mahieu, P. Ralle, B. Sédillot. Le modèle de microsimulation dynamique : DESTINIE. Département des Etudes Economiques d'Ensemble, Malakoff, 1999
- [**Burkhardt et al., 2004**] J. Burkhardt, S. Donikian, Y. Duthen, O. Héguy, D. Lourdeau, T. Morineau, M. Parenthoën, C. Sanza, J. Tisseau. Action spécifique humain virtuel : vers un humain synthétique temps-réel aussi vrai que nature. Rapport technique, CNRS, 2004
- [**Cornuéjols et al., 2002**] A. Cornuéjols, L. Miclet, Y. Kodratof. Apprentissage artificiel : Concepts et algorithmes. Editions Eyrolles, Paris, 2002
- [**Daguet, 1996**] F. Daguet. Mariage, divorce et union libre, Insee première n°482, 1996.
- [**Donikian, 2004**] S. Donikian. Modélisation, contrôle et animation d'agents virtuels évoluant dans des environnements informés et structurés. HDR de l'Université de Rennes1. URL : <http://www.irisa.fr/bibli/publi/habilitations/donikian/hdr.html> (consulté le 13.03.2007)
- [**Droesbeke et al., 2002**] J-J, J. Fine, G. Saporta. Méthodes bayésiennes en statistique. Editions TECHNIP, Paris, 2002
- [**Dupuis, 2002**] J. Dupuis. Estimation Bayésienne d'un modèle multi-état markovien. In Méthodes bayésiennes en statistique. Editions TECHNIP, Paris, 2002
- [**Frankhauser, 1991**] P. Frankhauser. Beschreibung der Evolution urbaner Systeme mit der Mastergleichung. Thèse de doctorat de l'Université de Stuttgart, 1991.
- [**Gibson, 1986**] J. Gibson. The ecological approach to visual perception. NJ : Lawrence Erlbaum Associates, Inc, Hillsdale, 1986
- [**Kerjosse et Tamby, 1996**] R. Kerjosse, I. Tamby. La situation démographique en 1994, mouvement de la population, Insee Résultats n°505-506, Paris, 1996
- [**Lamarche, 2003**] F. Lamarche. Humanoïdes virtuels, réaction et cognition : une architecture pour leur autonomie. Thèse de doctorat de l'Université de Rennes1, 2003
- [**Langlois, 2006**] P. Langlois. Approche conceptuelle de l'espace : structurations de l'espace, du temps et des objets dans un contexte de modélisation multi-agents. in Modélisation et simulation multi-agents : application pour les Sciences de l'Homme et de la Société, Editions Hermes-Lavoisier, Paris 2006
- [**Lefranc, 1995**] C. Lefranc. Le mariage en déclin, la vie en couple aussi, Insee première n°392, 1995.
- [**Lord et Levy, 1994**] R. Lord, P. Levy. Moving from cognition to action : A control theory perspective. In Applied psychology : an international review n°43, 1994
- [**Mallot, 1997**] A. Mallot. Behavior-oriented approaches to cognition : theoretical perspectives. Theory in biosciences n°116, 1997
- [**Mathis, 2003**] P. Mathis. Graphes et réseaux: modélisation multiniveau. Editions Hermès, Paris 2003
- [**Meyn et Tweedie, 1993**] S.P. Meyn, R.L. Tweedie. Markov Chains and Stochastic Stability.

New York Spring-Verlag, 1993

[**Naïm, et al., 2004**] P. Naïm , P-H. Wuillemain , P. Leray , O. Pourret , A. Becker. Réseaux bayésiens. Editions Eyrolles, Paris, 2004

[**Newell, 1990**] A. Newell. Unified Theories of Cognition. Harvard University Press, 1990

[**Pearl, 1986a**] J. Pearl. A constraint-propagation approach to probabilistic reasoning. In Uncertainty in Artificial Intelligence, Edition Elsevier Science, Amsterdam, 1986

[**Pearl, 1986b**] J. Pearl. Fusion, propagation, and structuring in belief networks. In Artificial Intelligence n°29, 1986

[**Pearl, 1987**] J. Pearl. Evidential reasoning using stochastic simulation of causal models. In Artificial Intelligence n°32, 1987

[**Pearl, 1987**] J. Pearl. Probabilistic reasoning in Intelligent Systems : Networks of Plausible Inference. Edition Morgan Kaufmann, San Francisco, 1988

[**Raftery et Tavaré, 1994**] A. E Raftery, S. Tavaré. Estimation and Modelling Repeted Patterns in High order Markov Chain with Mixture Transition Distribution Model. In Applied Statistics, vol. 43, 1994

[**Relieu et Quéré, 1998**] M. Relieu, L. Quéré. Mobilité, perception et sécurité dans les espaces publicsurbains. in Les Risques urbains : Acteurs systèmes de prévention. Collection VILLES, Paris, 1990

[**Robert, 1996**] C. Robert. Methodes de Monte-Carlo par chaînes de Markov. Economica, Paris, 1996

[**Robert-Bobée 2001**] I. Robert-Bobée. Modelling Demographic Behaviours in the French Microsimulation Model Destinie : an Analysis of Future Change in Completed Fertility, Insee document de travail n°G2001/14

[**Robert-Bobée 2002b**] I. Robert-Bobée. La programmation des événements démographiques dans Destinie, Insee note n°20/G211 26 juillet 2002

[**Saporta, 2006**] G. Saporta. Probabilité, analyse des données et statistique. Estimation, Monte-Carlo, Régression, Tests, Vraisemblance. Editions TECHNIP, Paris, 2006

[**Sardon, 1996**] J-P. Sardon. L'évolution du divorce en France. In Population, n°3, 1996

[**Thomas, 2005**] R. Thomas. Modèle de mémoire et de carte cognitive spatiales : application à la navigation du piéton en environnement urbain. Thèse de doctorat de l'Université de Rennes1, 2005

[**Toulemon, 1996**] L. Toulemon. La cohabitation hors mariage s'installe dans la durée. In Population n°3, mai-juin 1996

[**Vallin et Meslé, 2002**] J. Vallin, F. Meslé. Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle. Editions PUF, Paris, 2002

[**Villeneuve-Gokalp, 1994**] C. Villeneuve-Gokalp. Constance et inconstances de la famille, Travaux et Documents de l'Ined, 1994

