

CAHIER DU LAMSADE

293

Novembre 2009

**De l'aide multicritère à la décision
à l'aide multicritère à l'évaluation**

Un cadre et une application aux transports et à
l'environnement

De l'aide multicritère à la décision
à l'aide multicritère à l'évaluation

Un cadre et une application aux transports et à l'environnement

Benjamin Rousval^{*}, Denis Bouyssou[†]

Novembre 2009

^{*}rousval@free.fr, Lamsade, Université Paris Dauphine

[†]bouyssou@lamsade.dauphine.fr, Lamsade, Université Paris Dauphine

Résumé. Ce document vise à contribuer à l'élaboration d'outils d'aide multicritère à l'évaluation¹. A cet effet, est proposée une démarche qui consiste à structurer hiérarchiquement un ensemble d'objectifs (sous forme d'*arbre*), puis à appréhender leur atteinte à l'aide de *critères* dans un but d'établir un diagnostic, de donner l'alerte, de réaliser une analyse tendancielle et non systématiquement d'envisager une prise de décision. Cela nous amène naturellement à comparer l'*aide à la décision*, classique, à l'*aide à l'évaluation*. Ensuite, pour illustrer cette démarche dans le domaine des transports et de l'environnement, une structuration d'objectifs s'inspirant de la méthode proposée par Keeney² a été réalisée grâce à une série d'interviews. Puis, pour faire face au conflit qu'il existe entre la simplicité et la transparence du résultat de l'évaluation, nous proposons une problématique d'agrégation-désagrégation. Celle-ci consiste à réaliser, à tous les niveaux de l'arbre d'objectifs, une agrégation multicritère de type ELECTRE TRI³. Le cas de l'agrégation spatiale et temporelle est aussi traité. Nous donnons ensuite un exemple de modélisation rendant possible ces différentes agrégations. Cet exemple s'inspire des catégories de l'indice ATMO⁴ (indice bien connu dans le domaine de la qualité de l'air) tout en proposant une façon différente d'agréger les critères. A cette occasion, nous montrons qu'ELECTRE TRI est une extension naturelle d'ATMO.

Mots clés : Evaluation, Multicritère, Agrégation, Environnement, Transports.

Abstract. This document tries to contribute to the development of tools for multi-criteria aiding evaluation¹. For this purpose, it presents an approach which consists of structuring the problem through objectives, then to understand their reach by using *criteria* in order to establish a diagnosis, give warning, conduct a trend analysis and not systematically consider decision. It naturally leads us to compare *decision aiding* to *evaluation aiding*. Then, to illustrate this approach in transports and environment field, a series of interviews was conducted to structure objectives using a method strongly inspired by Keeney². Then, to face the conflict between the simplicity and the transparency of an evaluation's result, we propose an aggregation-desaggregation approach. This approach is to aggregate criteria at all the levels of the objectives tree, using ELECTRE TRI³ multicriteria aggregation method. We illustrate that by an example of modeling. This example uses categories of ATMO index⁴ (well-known index in the field of air quality) while proposing a different way of aggregating criteria. On this occasion, we show that ELECTRE TRI is a natural extension of ATMO.

Keywords : Evaluation, Multicriteria, Aggregation, Environment, Transports.

1. Rousval B., 2005, Aide multicritère à l'évaluation de l'impact des transports sur l'environnement, Thèse de doctorat, Université Paris Dauphine, Paris.

2. Keeney R.L., 1992, Value Focused Thinking - A Path to Creative Decisionmaking, Harvard University Press, Cambridge.

3. Roy B., Bouyssou D., 1993, Aide multicritère à la décision : méthodes et cas, Economica, Paris.

4. ADEME, 2000, Atmo 2000, Directive de 10 janvier 2000, Ministère de l'environnement, Paris.

Table des matières

Introduction	5
1 Processus d'aide à l'évaluation	7
1.1 Qu'est-ce qu'évaluer ?	7
1.1.1 Evaluation scolaire	8
1.1.2 Une définition de l'évaluation	10
1.2 Quelles sont les propriétés de l'évaluation ?	13
1.2.1 Modèle de pensée	13
1.2.2 Rationalité	14
1.2.3 Faits et Valeurs	15
1.2.4 Subjectivité ?	16
1.3 Pourquoi évaluer ?	18
1.4 Comment évaluer ?	19
1.4.1 Acteurs de l'évaluation	19
1.4.2 Processus d'aide à l'évaluation	20
1.5 Aide à la décision et aide à l'évaluation	22
1.5.1 Les problématiques de l'aide multicritère à la décision	22
1.5.2 Evaluer une décision	23
1.5.3 Différences avec l'aide à l'évaluation	24
2 Structuration des objectifs	26
2.1 Approche par les valeurs	26
2.1.1 Principes	26
2.1.2 Les objectifs	27
2.1.3 Les objectifs de fin et de moyen	28
2.1.4 Objectif stratégique	30
2.1.5 Révéler les objectifs	30
2.1.6 Construire l'arbre d'objectifs	31
2.1.7 Résultats : l'arbre des objectifs de fin	31
2.2 Interviews	33
2.2.1 Interviewés	33
2.2.2 Consigne de départ	34
2.2.3 Protocole	34
2.2.4 Préparation des interviews	34
2.2.5 Technique d'interviews	35
2.3 Résultats	36
2.3.1 Les différents arbres	36
2.3.2 Méthode de synthèse des arbres	37
2.3.3 Application	41
2.4 Retour d'expérience	43

3	L'évaluation et l'agrégation	44
3.1	Contexte de l'évaluation	44
3.2	Problématique de l'aide à l'évaluation	45
3.3	Agrégation multicritère	46
3.3.1	Méthodes d'agrégation utilisant un critère unique de synthèse . . .	46
3.3.2	Méthodes d'agrégation de type ELECTRE	47
3.4	Agrégation spatio-temporelle	48
3.5	Pondération des critères	50
3.5.1	Pondération des nœuds de l'arbre	51
3.5.2	Propagation des poids	51
3.5.3	Niveau de majorité	52
3.6	Exemple de modélisation	53
3.6.1	Le critère	53
3.6.2	Catégories	54
3.6.3	Seuils de discrimination	56
3.6.4	Seuil de veto	56
3.6.5	Principe d'agrégation spatio-temporelle	57
3.6.6	Le cas de l'indice Atmo	57
	Conclusion et perspectives	60
	Références	62
	Annexes	64

Introduction

Dans le contexte de la démocratie participative, une demande de plus en plus forte provenant des acteurs (les élus ou responsables politiques, l'opinion publique et les représentants associatifs) témoigne du besoin d'outils permettant d'appréhender au mieux l'impact des transports sur l'environnement. Ces outils doivent permettre de considérer les aspects environnementaux conjointement aux aspects économiques et sociaux afin de contribuer à un développement durable des transports. Répondre à cette demande est une tâche très ambitieuse qui a donné, entre autre, naissance à un projet de recherche, le projet « Prospectives et Indicateurs Environnementaux » (PIE) [21, 16] mené par le Laboratoire Transports et Environnement de l'Institut National de Recherche sur les Transports et leur Sécurité (Inrets-Lte). Il s'agit avant tout d'une demande d'instrumentation afin d'aider à évaluer, de façon globale, l'impact des transports sur l'environnement. Ce document s'inscrit dans la continuité de la thèse de Rousval [22]. Il présente et complète certains résultats grâce à trois parties pouvant être considérées comme indépendantes même si elles contribuent à répondre à une même question de recherche bien spécifique : celle de l'aide multicritère à l'évaluation de l'impact des transports sur l'environnement.

Cela nous amènera tout naturellement à nous interroger sur ce qu'est l'évaluation et plus précisément l'aide à l'évaluation (section 1). Dans ce contexte, faute d'actions potentielles (puisqu'il s'agit d'évaluer un système et non de comparer différentes options) nous proposerons une approche fondée sur la structuration d'objectifs qui est une façon de modéliser une partie d'un système de valeurs, dans le but de servir de référentiel pour l'évaluation. Nous comparerons cette approche aux méthodes d'aide à la décision.

Nous présenterons en section 2 une méthode permettant de construire un arbre d'objectifs, méthode proposée par Keeney [14] qui distingue les objectifs de fins des objectifs de moyens. Cette démarche sera illustrée par le résultat d'une série d'interviews ayant été menés auprès de décideurs dans le domaine de l'environnement et des transports.

Dans le domaine des transports et de l'environnement, les décideurs sont bien souvent confrontés à l'analyse de longs rapports d'experts mélangeant à la fois des informations descriptives, qualitatives et quantitatives. C'est l'une des raisons à l'origine de la demande d'instrumentation dont il est question dans le projet PIE. Ces décideurs ont en effet une grande difficulté, d'une part à considérer chaque aspect dont il est question dans les rapports d'experts, d'autre part à se forger une opinion globale. L'intelligibilité et l'agrégation de l'information sont donc souhaitées. Notons de plus que les processus de décisions publiques font souvent intervenir des groupes d'acteurs dont les enjeux ou valeurs ne sont sensiblement pas les mêmes. Or, l'activité d'aide à l'évaluation telle que nous l'envisageons ici utilise les valeurs des acteurs comme référentiel de l'évaluation. Dans un contexte multi-décideur, il n'y a pas un mais plusieurs référentiels a priori différents. Cela complique donc la tâche et nous incite notamment à soulever la questions du « comment intégrer différents systèmes de valeurs dans une démarche d'évaluation ? ». La section 3 est donc consacrée au problème de l'agrégation du résultat de l'évaluation quand cette évaluation présente de nombreuses composantes, de surcroît hétérogènes, dans un contexte multi-acteurs. On y propose certaines orientation méthodologiques

illustrées par la modélisation d'un critère [22]. Enfin, nous y présentons une mise en parallèle (du point de vue de leur principe d'agrégation) entre, d'une part, l'indice Atmo [2], bien connu dans le domaine de la santé et de la qualité de l'air, et d'autre part, la méthode ELECTRE TRI [25].

1 Processus d'aide à l'évaluation

La préoccupation centrale de cette première partie est de proposer démarche pour évaluer un système complexe. Par exemple, comment évaluer l'impact environnemental des transports sur une ville ? Evidemment cette tâche n'est pas aisée tant les transports ont des impacts divers et variés (pollution locale, pollution globale, bruit, impact sur le paysage, différence d'échelles temporelles et géographiques d'observation de ces impacts...). Elle donne naissance à un problème multidimensionnel ou multicritère. De ce fait, nous parlerons donc d'« aide à l'évaluation » : il s'agit alors d'aider les acteurs concernés à se forger une opinion sur l'état de la situation de la ville considérée, d'un système complexe.

Pour mieux cerner l'évaluation et tenter de définir ce que nous entendons par « aider à évaluer », nous proposons tout d'abord d'essayer de répondre aux questions suivantes :

- qu'est-ce qu'évaluer ?
- quelles sont les propriétés de l'évaluation ?
- pourquoi évaluer ?
- comment évaluer ?

A partir de ces éléments de réflexion, nous proposerons une façon d'envisager l'aide à l'évaluation grâce à la description d'un processus et de ses différentes étapes. Enfin, nous rappellerons les différentes problématiques de l'aide à la décision et nous verrons en quoi elles sont utilisables en aide à l'évaluation.

1.1 Qu'est-ce qu'évaluer ?

Devant la diversité des activités d'évaluation, nous proposons de nous référer tout d'abord aux définitions données par le dictionnaire *Petit Robert* :

- « Porter un jugement sur la valeur de »,
- « Calculer, chiffrer, fixer approximativement, estimer, apprécier »,
- « Déterminer une quantité par le calcul, sans recourir à la mesure directe ».

On découvre ici les multiples facettes de l'évaluation. Selon ces définitions, on s'aperçoit que l'évaluation peut très bien être fondée sur le jugement d'une ou plusieurs personnes concernant la valeur d'une chose, son appréciation. Mais elle peut tout aussi bien être fondée sur une mesure ou le un calcul (comme pour une longueur par exemple). Remarquons cependant que, la mesure, le calcul peut être très précis et univoque ou bien il peut s'agir d'une approximation, d'une estimation. Nous sommes donc face à une activité variée alliant jugements de valeurs et observations plus ou moins univoques ou précises. Dans la littérature, plusieurs domaines traitent de l'évaluation, que se soit implicitement ou explicitement. Nous souhaitons dans un premier temps nous référer au domaine scolaire pour deux raisons. D'une part, car depuis fort longtemps l'évaluation est une question importante dans le scolaire. D'autre part, car cette question centrale a donné lieu à une littérature bien spécifique.

1.1.1 Evaluation scolaire

Une des problématiques des systèmes scolaires est l'évaluation des élèves. Cela peut se traduire par l'évaluation d'une production (une copie par exemple) ou d'un ensemble de productions réalisées par un élève. Regardons comment est alors définie l'évaluation :

« Dans son acception la plus large, le terme évaluation désigne l'acte par lequel, à propos d'un événement, d'un individu ou d'un objet, on émet un jugement en se référant à un (ou plusieurs) critère(s), quels que soient par ailleurs ce(s) critère(s) et l'objet du jugement. » Noizet [20]

Dans cette définition, Noizet ne restreint pas l'évaluation à une activité scolaire. On y voit apparaître plusieurs notions. La première notion est ce sur quoi porte l'évaluation. Autrement dit, ce qui répond à la question « qu'évalue-t-on ? ». Choisissons d'appeler « évalué », ce qui est évalué, ce qui fait l'objet de l'évaluation. La seconde notion concerne l'individu ou l'institution qui évalue, qui définit et ordonne l'évaluation, qui est intéressé directement par le résultat de l'évaluation : nous la nommerons *évaluant*. Revenons à la définition de Noizet. Pour l'auteur, l'évalué est soit un événement, soit un individu, soit un objet. L'évaluant, lui, n'est précisé qu'au travers du « on ». En cela, Noizet ne restreint pas l'auteur de l'activité à une seule personne, et permet d'envisager l'évaluant comme un groupe de personnes, ou toute autre chose (une institution). L'activité d'évaluation est définie par Noizet comme le fait d'émettre un jugement en se référant à un (ou plusieurs) critère(s). Le fait d'émettre un jugement coïncide bien avec une des définitions du dictionnaire. Selon Noizet, le moyen permettant d'émettre ce jugement est de se référer à un ou plusieurs critères, c'est-à-dire que le jugement consiste finalement à retenir le ou les critères jugés pertinents par l'évaluant pour réaliser sa propre évaluation. Il lui confère la liberté du choix des critères en ajoutant « quels que soient par ailleurs ce(s) critère(s) ». Les critères retenus n'ont pas forcément besoin d'être justifiés puisqu'il s'agit là d'un jugement de valeur propre à l'évaluant. Enfin, il le laisse aussi maître de choisir l'objet du jugement. Il semble que ce qu'il nomme l'objet du jugement est la raison d'être de l'évaluation, autrement dit, ce qui répond à la question « pourquoi évaluer ? ».

Intéressons nous maintenant à une seconde définition de l'évaluation qui nous est donnée par Abernot [1] : « L'évaluation consiste en une mesure ou une appréciation, à l'aide de critères, de l'atteinte d'objectifs ou du degré de proximité d'une production par rapport à une norme. » Nous allons tenter de voir en quoi cette définition rejoint celle de Noizet, en quoi elle s'en écarte et en quoi elle la complète. Tout d'abord, tout comme Noizet, Abernot ne fait pas allusion à l'évaluant. Il ne nous informe pas sur « qui » réalise l'évaluation. Cela laisse toute ouverture sur la nature de l'évaluant. C'est donc ici un point commun. Comme Noizet, Abernot nous parle de critères. Il semble que tout deux accordent à ce mot la même signification, mais Abernot apporte une précision supplémentaire. En effet, pour lui, les critères permettent d'apprécier l'atteinte d'objectifs. Ce qui sous-entend que les objectifs de l'évaluant doivent être explicités par une étape amont. Les critères servent ensuite à informer sur l'atteinte de ces objectifs. C'est une différence majeure. En effet, la définition de Noizet laisse l'évaluant libre de choisir ces critères en fonction de son propre système de valeurs sans pour autant avoir besoin d'explicitier les motivations, les objectifs sous-jacents qui sont à l'origine de ces

choix. Autrement dit, la définition d'Abernot semble imposer une étape d'explicitation du système de valeurs de l'évaluant par le biais d'objectifs. Cela impose une formalisation qui, par nature, ajoute de la transparence quant aux motivations concernant les choix des critères qui serviront à l'évaluation. Cette différence est à notre sens fondamentale. Noizet laisse l'évaluant libre de choisir ses critères sans formaliser les objectifs. Pour Abernot, l'évaluant émet son jugement de valeur grâce à des objectifs et non grâce à des critères. Il s'agit ensuite de choisir les critères permettant d'apprécier l'atteinte de ces objectifs. On peut imaginer que ce travail de construction de critères peut relever de la compétence d'un expert assistant l'évaluant. Ce dernier pourra alors pratiquer l'évaluation dans un domaine qu'il ne maîtrise pas entièrement. En effet, s'il n'a pas besoin de choisir les critères d'évaluation, mais simplement de décliner ses objectifs, il peut néanmoins conserver toute sa légitimité d'évaluant. Alors que la définition de Noizet implique que l'évaluant sait choisir, sait interpréter les critères retenus. Il sait « se référer à » ces critères pour « émettre son jugement ». L'avantage de l'approche par les objectifs semble donc double : d'une part elle contribue à plus de transparence et d'autre part elle offre la possibilité pour l'évaluant de ne pas forcément posséder une expertise du domaine concerné. Notons ensuite que concernant l'évalué, Abernot a choisi le mot « production ». On sent bien que ce mot rappelle le domaine scolaire. Notons tout de même qu'évaluer la production n'est pas la finalité. L'évalué (ce qui est évalué au travers des productions) est bien l'élève, l'individu. Les productions traduisent, reflètent les caractéristiques de l'évalué (car c'est lui qui les produit). Ce ne sont pourtant pas elles que l'on cherche à juger, mais bien l'élève. Les objectifs que l'évaluant possède dans son système de valeurs concernent bien entendu l'évalué. Abernot fait ensuite référence à une *norme*. L'évaluation consiste pour lui à mesurer (ou apprécier) « le degré de proximité d'une production à une norme ». Existe-t-il une différence fondamentale entre un critère et une norme ? En fait, il semble qu'une norme reflète l'atteinte d'un objectif commun à plusieurs systèmes de valeurs, ou plus précisément aux systèmes de valeurs de plusieurs évaluateurs. Elle est souvent le résultat d'un consensus entre plusieurs évaluateurs concernant l'objectif qu'elle reflète (exemple : une norme européenne devient commune à différents pays). La norme permet alors la comparaison du résultat d'évaluations réalisées par différents évaluateurs. Dans sa définition Abernot distingue mesurer et apprécier. Revenons sur la différence entre ces deux notions. Comme nous l'avons vu précédemment, on peut considérer que la mesure est le résultat d'un processus objectif tandis que l'appréciation est un jugement de valeur. L'auteur nous laisse donc la liberté d'appréhender l'atteinte des objectifs de différentes manières. Par exemple, la correction d'un questionnaire à choix multiple aboutira toujours à un même résultat quel que soit le correcteur (si le barème est défini de façon très précise). Elle relève donc d'une démarche relativement objective, tandis que la correction d'une copie littéraire relève beaucoup plus d'une appréciation, d'un jugement de valeur. Idéalement, on souhaite pourtant qu'une copie littéraire soit appréciée de la même façon par tous les enseignants, que la note de l'élève ne dépende pas de la personne qui la corrige. Mais on sait bien que cela reste malheureusement très difficile. Finalement, dans les deux cas, l'objectivité est souhaitée mais elle est plus aisée à atteindre dans le cas du QCM que dans le cas de la production littéraire. Abernot autorise

donc une certaine souplesse dans la façon d'évaluer car il est conscient que l'objectivité n'est pas toujours possible même si elle est souhaitée autant que faire ce peut.

Étudions maintenant la définition de De Ketele [10]. Pour l'auteur, évaluer signifie :

- « recueillir un ensemble d'informations suffisamment pertinentes, valides et fiables,
- et examiner le degré d'adéquation entre cet ensemble d'informations et un ensemble de critères adéquats aux objectifs fixés au départ ou ajustés en cours de route en vue de prendre une décision. »

Cette dernière définition est la plus complète des trois. Elle précise que l'activité intègre une étape de recherche d'information servant à la construction des critères représentatifs d'objectifs. De Ketele ajoute aussi que les objectifs fixés a priori peuvent être ajustés au cours du temps. Cela enrichit l'activité d'évaluation et lui confère un caractère itératif. Autrement dit, on n'évalue pas une fois pour toute, mais on ajuste l'évaluation au cours du temps. On peut considérer alors l'évaluation comme un processus alternant des phases de définition d'objectifs et de recueil d'informations. L'auteur précise ici l'objet de l'évaluation alors que Noizet laissait une grande liberté sur cet objet (« quel que soit l'objet du jugement »). Pour De Ketele il s'agit avant tout de prendre une décision (« en vue de prendre une décision »). Pour cet auteur, identifier la prise de décision est un pré-requis qui permet de distinguer l'évaluation de la simple appréciation, de la recherche d'information. Dans le sens de la définition d'Abernot, les décisions peuvent être par exemple :

- d'accepter le passage d'un élève à un niveau supérieur,
- de délivrer un diplôme,
- de recueillir les meilleurs éléments d'une promotion (principe du concours),
- d'ajuster le contenu du cours pour les années suivantes

Nous pensons qu'il est tout à fait raisonnable d'envisager l'évaluation scolaire comme une activité forcément liée à une prise de décision. Nous désirons, pour notre part, élargir le champ de l'évaluation. En effet, nous souhaitons pouvoir évaluer un système complexe, simplement pour informer ou communiquer sur l'état de ce système. Dans ce cas, nous faisons effectivement de la recherche d'information, de l'appréciation, mais nous souhaitons aussi émettre un jugement de valeur sur l'état du système. A notre sens, il s'agit donc aussi d'une activité d'évaluation, même si aucune prise de décision n'est a priori identifiée. Nous pensons qu'un tel type d'évaluation est d'ailleurs susceptible de faire naître de nouveaux enjeux décisionnels, qui n'auraient pas pu être identifiés auparavant.

Nous reviendrons sur les différents buts que nous accordons à l'évaluation dans le paragraphe suivant, mais nous ne souhaitons pas restreindre ici notre définition de l'évaluation à quelque chose de strictement dépendant d'une prise de décision. En ce sens nous sommes plus proche de la vision des choses de Noizet qui, lui aussi, choisi de généraliser la portée de sa définition en s'éloignant du cadre de l'évaluation scolaire.

1.1.2 Une définition de l'évaluation

L'évaluation est un processus qui vise à quantifier et/ou qualifier un système grâce à toute information nécessaire à la construction de critères permettant d'appréhender

au mieux l'atteinte de l'ensemble des objectifs concernant ce système et jugés pertinents dans le cadre d'une activité plus large mais identifiée au préalable.

Nous allons commenter cette définition point par point.

- « viser à qualifier et/ou quantifier un système » : nous avons choisi d'utiliser le terme « système » pour laisser plus de liberté concernant la nature de l'évalué. Nous pensons notamment que l'évaluation peut s'appliquer à des systèmes plus complexes qu'un individu (qui l'est déjà très!), qu'une chose ou qu'un événement. Dans le cadre de l'évaluation environnementale par exemple, l'évalué pourrait être une ville, une partie de territoire, un ensemble géographique, ce qui s'apparente bien plus à un système qu'à une chose. Nous « visons à qualifier et/ou quantifier ce système ». En effet, il nous semble que le résultat de l'évaluation pourrait être une ou plusieurs quantifications, une ou plusieurs qualifications, ou une composition des deux. Nous avons préféré les termes « qualifier » et « quantifier » aux termes « mesurer » et « apprécier ». En effet, une mesure est quelque chose de très précis et rigoureux or il n'est pas toujours possible de l'obtenir selon la nature de ce qui est quantifié : si mesurer, c'est avant tout quantifier, l'inverse n'est pas vrai. Il est possible d'obtenir des valeurs numériques, des quantités, qui n'ont pas toutes même les propriétés. Prenons par exemple, les relevés de température atmosphérique. S'il fait 20°C à Marseille et 10°C à Paris peut-on affirmer qu'il fait deux fois plus chaud à Marseille qu'à Paris ? Non, car ce rapport n'est plus vrai si l'on relève les températures en Fahrenheit, si l'on change de repère. L'appréhension de la température se fait relativement à un référentiel mais le zéro de ce référentiel est conventionnel (température de solidification de l'eau). Zéro euro représente l'absence de monnaie, tandis que 0°C ne représente pas l'absence de température. De ce fait, on peut dire qu'un vêtement à 20 euros est deux fois plus cher qu'un autre à 10 euros tandis qu'on ne peut pas dire que 20°C est deux fois plus chaud que 10°C. Ainsi, les rapports entre valeurs ne sont pas toujours pertinents. Il faut donc faire attention à la nature, aux propriétés et au sens que l'on accorde aux nombres que l'on manipule. En évaluation scolaire, on considère parfois des valeurs numériques dont la nature peut être plus ou moins pauvre d'un point de vue de leur signification. Par exemple, la note 20/20 est-elle deux fois meilleure que la note 10/20 ? Ou encore, la note 1/20 est-elle infiniment meilleure que la note 0/20 ? Ou bien, passer de 9/20 à 11/20 est-il équivalent que de passer de 14/20 à 16/20 ? Dans d'autres cas, il est possible de considérer des évaluations non numériques comme des qualifications. Le terme « apprécier » englobe dans sa signification un certain degré de jugement de valeur. Nous pensons que le jugement de valeur est déjà présent dans notre définition lorsqu'il s'agit de préciser les « objectifs jugés pertinents » puis de construire les critères permettant de les représenter. Selon nous, l'évaluateur choisit les objectifs pertinents selon ses propres valeurs et c'est en cela qu'il permet le « jugement ». Le terme « qualification » ne concerne que la nature de l'information considérée : elle peut en effet être un simple qualificatif (en opposition à une quantité) concernant le système. Cette qualification peut tout à fait être établie de manière univoque (par exemple, une couleur) tandis que

l'appréciation est fondée sur le jugement d'une ou plusieurs personnes.

- « grâce à toute information nécessaire à la construction de critères » : l'évaluation est basée sur le recueil d'un ensemble d'informations. Il est possible qu'une information brute soit représentative pour appréhender l'atteinte d'un objectif. Cependant, les informations collectées peuvent souvent être entachées d'imprécisions. De plus, ces informations brutes ne sont pas forcément suffisamment représentatives d'un objectif. Pour ces deux raisons, les informations doivent souvent être traitées afin de construire des critères. Il s'agit de transformer, de regrouper, de réaliser des opérations sur ces informations brutes afin d'obtenir un critère le plus représentatif possible du point de vue ou de l'objectif qu'il est censé représenter. La différence entre l'information brute et le critère est donc le lien avec l'objectif. Le critère est représentatif alors que l'information brute, pas forcément. Le critère est finalement l'outil retenu pour repérer la satisfaction d'un objectif, d'un point de vue.
- « permettant d'appréhender au mieux l'atteinte de l'ensemble des objectifs » : pour appréhender l'atteinte d'un objectif il est nécessaire de construire un critère censé le représenter au mieux. Afin d'évaluer un système de manière globale, il faut appréhender l'atteinte de l'ensemble des objectifs retenus par l'évaluant. Pourquoi « au mieux » ? D'une part car il s'agit d'abord de choisir le critère qui va représenter l'objectif, et ce choix est une prise de position parmi d'autres. D'autre part, car la construction du critère relève souvent de la connaissance d'experts et celle-ci n'est pas forcément parfaite : elle peut être entachée d'incertitude, d'imprécision ou être incomplète selon le domaine concerné.
- « concernant le système et jugés pertinents » : les objectifs que l'évaluant se doit de décliner concernent l'évalué, c'est-à-dire un système. L'évaluant est seul juge des objectifs qu'il désire retenir, de ceux qu'il considère pertinents. Il va choisir de retenir ou d'écarter des objectifs selon son propre jugement de valeurs. Ce choix dépend à nouveau d'une ou plusieurs personnes, d'un évaluant.
- « dans le cadre d'une activité plus large mais identifiée au préalable » : tout le processus d'évaluation est réalisé dans un but précis. On n'évalue pas pour rien. Or, si l'on évalue dans un but que l'on a mal identifié dès le début, il y a fort à croire que l'on risque de pratiquer une évaluation en décalage avec nos intentions premières. En conséquence, le processus d'évaluation risque d'être mal cadré, et le résultat de l'évaluation s'en retrouvera affecté.

Pourquoi voir l'évaluation comme un *processus*? Selon notre définition, on voit apparaître un ensemble de tâches qui constitue l'activité d'évaluation : définir le but de l'évaluation, définir ce qui est évalué, définir celui qui évalue, définir les objectifs, etc. De plus, cette activité nécessite l'intervention de différents acteurs : évaluant, experts, évalué éventuellement, personnes chargées de la collecte de l'information, etc. Un troisième point tend à nous faire pencher pour le terme processus d'évaluation. Il s'agit du fait que les objectifs de l'évaluant concernant l'évalué sont susceptibles de changer au cours du temps. Cela implique que certaines tâches devront être réalisées plusieurs fois.

1.2 Quelles sont les propriétés de l'évaluation ?

1.2.1 Modèle de pensée

D'après Vial [28], « Les modèles de pensée nous agissent, ils nous donnent des cadres pour concevoir l'être-au-monde. Ils ne sont pas propre à l'évaluation, mais supervisent tous les domaines des sciences humaines ». Les modèles de pensée présentés par l'auteur sont les suivants :

- Le déterminisme pose que les éléments sont des substances, des essences liées par la causalité. Proche du fatum, inspirée par la philosophie fataliste, du destin, elle est d'abord dans la causalité mono-linéaire (modèle de l'explication causale, évaluation des effets), puis dans la causalité pluricausale mais toujours linéaire (dite différentielle).
- Le fonctionnalisme se situe dans le schéma moyens-fins. Eminemment représenté par la théorie des objectifs issue de la polémologie, avec ces cibles à détruire au nom de l'efficacité, il se donne aussi à voir dans la prise de décision rationnelle et la résolution de problème. Il s'adresse à un Homme rationnel, pour la rationalisation des pratiques.
- Le structuralisme avec sa mise à jour d'invariants, voire d'universaux. Les éléments deviennent des virtualités reliées par des rapports stables dans une structure, elle, virtuelle, résultat d'une lecture analytique ; accomplissant des transformations (nommées « fonctions que la structure remplit ») et s'actualisant, par une série de variations de surface, dans les pratiques sociales.
- La systémie⁵, ne s'occupe plus de la nature des éléments mais de leurs interrelations. L'élément est l'ensemble des relations qu'il entretient aux autres éléments auxquels il est connecté. Les différents systèmes se différencient par leur plus ou moins grande ouverture mais sont toujours au service d'une fonction puis de plusieurs fonctions à remplir.
- La complexité, est un modèle fort ancien autrefois réservé aux cercles initiatiques, une gymnastique de l'esprit pour concevoir le sujet-au-monde. En tant que modèle de recherche et d'évaluation, il est en cours d'élaboration.

Face à cela, l'auteur propose ensuite une classification des modèles d'évaluation en trois catégories :

- l'évaluation-mesure (modèles historiques : explication causale, métrie, docimologie), provenant du déterminisme.
- l'évaluation-gestion (modèles contemporains : rationalisation par les objectifs, évaluation comme aide à la prise de décision, cybernétiques, systémie), découlant du fonctionnalisme, du structuralisme et de la systémie,
- l'évaluation-problématisation du sens (modèle en élaboration : systémique, dialectique, herméneutique), issue de la systémie, de la dialectique contemporaine de la psychanalyse et de l'herméneutique.

Selon les catégories de l'auteur, le présent travail prend place dans l'évaluation-gestion. L'évaluation est en effet être une approche de rationalisation par les objectifs.

5. plus souvent appelée systémique

La distinction fin-moyens semble être une voie intéressante pour structurer les objectifs de l'évaluant. La définition retenue est donc rattachée au fonctionnalisme. Et comme le précise Vial, elle « s'adresse à un Homme rationnel, pour la rationalisation des pratiques ». Cependant, nous souhaitons préciser que cette rationalisation doit être prise dans le contexte d'un homme en action et en débat (contexte multi-acteurs dont les enjeux, les objectifs peuvent différer, voire diverger).

1.2.2 Rationalité

Dans un contexte multi-acteurs et face à la complexité de la tâche d'évaluation, il semble important de se demander vers quel type de rationalité il est souhaitable de tendre.

Contrairement à Weber qui voyait en la multiplicité des formes de rationalité l'impossibilité d'une quelconque validation, Habermas réhabilite la place de la rationalité dans la société en spécifiant les différentes formes qu'elle peut prendre et en l'accompagnant d'une variété de formes de validations [18]. Habermas [9] précise à cette occasion que, dans la société moderne, la « vérité scientifique » est la forme de rationalité qui domine les autres. Il propose alors de rééquilibrer les choses en laissant plus de place à trois autres formes de rationalité. Ainsi, selon l'auteur, ce rééquilibrage aboutit à une meta-rationalité appelée « rationalité communicative », qui, pour prétendre à la validité, doit permettre la confrontation et la cohabitation des quatre formes de rationalité suivantes :

- L'intelligibilité : les acteurs doivent utiliser un langage compréhensible pour formuler leurs messages. Selon l'auteur, cette première forme de rationalité est une condition prérequis aux trois autres formes.
- La vérité scientifique : chaque message doit être conforme aux théories scientifiques en vigueur. Habermas ajoute que cette vérité n'est pas immuable ; elle suit l'évolution des théories scientifiques.
- La justesse normative : le message est valide s'il fait appel à un ensemble de normes et de valeurs auxquels les acteurs adhèrent. Par exemple, « voler est mal » est un message valide si tous les acteurs partagent cette valeur.
- La sincérité : Le message est sincère s'il y a adéquation entre ce qui est dit dans le message et ce qui est pensé par l'émetteur du message.

La validité d'un message est liée à sa capacité à satisfaire ces différentes formes de rationalités. Or, le dialogue et la concertation entre acteurs semblent nécessaires pour tendre vers celles-ci (la « vérité scientifique » nécessite la mise en place d'un dialogue avec la communauté scientifique, la « justesse normative » nécessite une confrontation des valeurs des différents acteurs). Dès lors, la « rationalité communicative » de Habermas plaide en faveur de la mise en place de démarches participatives.

Dans le processus d'évaluation, chacun des acteurs sera amené à être rationnel non pas uniquement par rapport à ses convictions, mais aussi par rapport à la situation de négociation, de recherche de compromis que le processus engendre.

1.2.3 Faits et Valeurs

Nous souhaitons faire la distinction dans ce paragraphe entre ce qui relève des *faits* et ce qui relève des *valeurs*.

Nous entendons par *fait*, l'ensemble des observations, des données, des informations, des connaissances de phénomènes, des productions, qui vont servir à l'évaluation d'un système. En contexte multi-acteurs, il semble y avoir deux types de faits :

- les faits partagés,
- les faits potentiellement partagés.

Les « faits partagés » sont les faits pour lesquels il existe un réel consensus concernant leur fiabilité, leur crédibilité, leur précision. Par exemple, un relevé de température Atmosphérique réalisé selon un protocole connu (sous abri, à une heure bien précise) est une donnée pour laquelle il existe un certain consensus concernant sa fiabilité, sa précision et sa signification. On peut difficilement remettre en cause une telle information. Il s'agit d'un fait partagé.

Un « fait potentiellement partagé » est au contraire un fait qui ne fait pas l'unanimité. On peut par exemple considérer une note attribuée à une production littéraire par un correcteur. Pour de nombreuses raisons, il est possible que la note obtenue par l'élève avec ce correcteur ayant ses propres sensibilités, appartenant à tel groupe de copie, placé à tel position dans la pile, dans le contexte bien précis de l'interrogation, ait été différente dans d'autres conditions. Or ces conditions ne sont pas forcément explicitées, ni prises en compte lors de l'appréciation du résultat de l'élève. Ce fait est donc de nature différente de celle du relevé de température. Sa fiabilité et son degré de précision sont tout à fait discutables. En ce sens, ce fait est potentiellement partagé.

C'est ensuite en fonction de son propre système de valeurs qu'un acteur peut émettre un jugement concernant un fait. Ce jugement semble présenter trois niveaux :

- la crédibilité du fait,
- la pertinence ou représentativité du fait,
- l'importance que l'on confère à ce qu'il représente.

Il semble assez raisonnable de considérer les « faits partagés » comme crédibles pour tous les acteurs. Cependant, un fait est rarement complètement irréfutable. Il est possible qu'une personne considère un « fait partagé » comme « peu crédible » même s'il va ainsi à l'encontre d'une certaine unanimité. Ceci est évidemment encore plus fréquent quand il s'agit de « faits potentiellement partagés », pour lesquels les arguments de réfutation sont plus aisés à avancer. On peut par exemple, remettre en question la crédibilité de la note d'une production littéraire, notamment pour les raisons évoquées précédemment.

Ensuite, on peut aussi remettre en cause la représentativité d'un fait. On peut par exemple croire qu'une note de philosophie, aussi crédible qu'elle soit, n'est pas forcément représentative des aptitudes d'un élève dans ce domaine. En effet, on peut juger peu crédible d'appréhender le niveau de philosophie d'un élève sur une simple note. C'est le choix du fait en tant que représentant de la valeur qui est ici remis en question.

Enfin, l'acteur tient compte de l'importance qu'il confère au fait, et cela sans considération de la crédibilité qu'on lui accorde. Une personne peut en effet juger important le fait d'avoir des bonnes notes en philosophie, et cela dans l'absolu. Ces considérations

n'ont rien à voir avec la crédibilité que l'on accorde à la notation d'une production, mais bien à la valeur, l'importance que l'on accorde à ce que représente la philosophie en elle-même.

Voici un exemple illustrant ces propos. Une personne souhaite évaluer le climat d'une ville qu'elle voudrait visiter. Elle dispose pour cela, d'un relevé de température effectué quelques jours auparavant à midi heure locale. Elle peut considérer ce relevé, ce fait, comme peu fiable car elle ne connaît pas toutes les conditions dans lesquelles il a été réalisé (sous abri ou non). Il s'agit donc d'un fait potentiellement partagé. Cependant, elle le juge crédible, car elle pense qu'il y a peu d'écart entre les relevés sous abri et les autres. En revanche, elle trouve qu'un seul relevé n'est peut-être pas très représentatif du climat de la ville (on néglige alors le taux d'humidité, etc). Cela ne l'empêche pas d'apporter une très grande importance au climat des villes qu'elle visite.

Cette personne va donc se constituer sa propre vision concernant les relevés de températures dont elle dispose en tenant compte de :

- la crédibilité : ici assez forte,
- la représentativité : ici assez faible,
- l'importance : ici très forte.

Un des intérêts de concevoir l'évaluation comme un processus est de permettre l'itération entre des étapes de définitions et déclinaisons de valeurs et des étapes de sélections et collectes de faits. Cette distinction entre faits et valeurs semble apporter de la transparence à l'évaluation en permettant d'identifier plus facilement les rôles des différents acteurs du processus.

1.2.4 Subjectivité ?

Dans la conclusion d'un article, Gérard [13] met en garde en parlant de la subjectivité de l'évaluation (scolaire). « La subjectivité est donc inévitable dans tout processus d'évaluation, mais elle est aussi indispensable. Subjectivité, oui, mais arbitraire, non ! »

Nous allons étudier le point de vue de Gérard. Il part de la définition de De Ketele [10] qui fait référence dans le domaine et que nous avons citée précédemment. Pour Gérard, dans le processus d'évaluation, les cinq sources suivantes d' « inévitable et indispensable subjectivité » sont présentes :

- « le choix du type de décision et de l'objectif de l'évaluation » : n'oublions pas que pour De Ketele, l'évaluation prépare une prise de décision. L'auteur écrit « qu'il y a toujours quelque part quelqu'un qui choisit le type d'objectif que doit poursuivre l'évaluation et que ce choix va orienter le processus d'évaluation. Première inévitable et indispensable subjectivité ». Il paraît tout à fait raisonnable de partager ce point de vue. En effet, s'il n'y a pas de motivation à l'évaluation, l'évaluation n'a pas lieu d'être. Si à l'inverse, l'évaluant veut évaluer pour telle ou telle raison, tel ou tel but, il fait un choix qui conditionne évidemment le processus d'évaluation. Mais l'évaluant est l'acteur directement intéressé par le résultat de l'évaluation et ses motivations sont l'unique raison d'être de l'évaluation. Autrement dit, s'il l'on désire évaluer, c'est que l'on a une bonne raison de le faire. Si l'on n'a pas de bonnes

- raisons alors il n’y a pas à évaluer. Et cette « bonne raison » est effectivement personnelle à l’évaluant.
- « le choix des critères » : l’évaluant choisit lui-même les critères qui vont servir à évaluer. Bien sur, ceci relève directement du système de valeurs de l’évaluant. Dans notre approche, ce choix est celui des objectifs. L’évaluant, guidé par son système de valeurs, juge en choisissant les objectifs qu’il désire prendre en compte pour évaluer. Cette étape est inévitable car sans objectifs, il n’y a pas de critère d’évaluation et donc il n’est pas possible d’évaluer. A l’inverse, prendre en compte de façon exhaustive tous les objectifs possibles et imaginables paraît difficile et même dangereux selon Cardinet [6] : « ... pour y arriver (à l’objectivité de l’évaluation), il faudrait une exhaustivité des objectifs, des critères, des indicateurs, ... Les conséquences sociales seraient dès lors inacceptables, car une telle évaluation déboucherait sur une standardisation extrême ». Donc le choix des objectifs par l’évaluant est un « jugement de valeur » inévitable mais aussi indispensable au processus d’évaluation.
 - « le choix des indicateurs » : il entend par indicateurs, « les faits, que l’on peut constater, et les représentations, c’est-à-dire ce que les personnes pensent, ce qu’elles disent ». Effectivement, le choix des données que l’on va observer est une prise de position parmi d’autres. Ce choix dépend évidemment de l’accessibilité des données mais aussi de la signification que l’on accorde à la donnée observée. Cette signification peut bien souvent émaner d’un avis d’expert. Il s’agit alors effectivement d’un jugement mais qui n’est cependant plus du ressort de l’évaluant. Parfois au contraire, cette signification peut s’imposer d’elle même et en conséquence ne pas dépendre d’un jugement individuel. C’est le cas notamment lorsque la signification d’un indicateur est bien établie, lorsqu’elle relève de la « connaissance partagée ». Il y a alors dans le choix de l’indicateur beaucoup moins de « subjectivité » au sens de l’auteur.
 - « le choix de la stratégie » : Gérard entend par stratégie, la méthode retenue pour la recherche de l’information. C’est en effet très caractéristique du domaine scolaire de s’interroger sur la méthode de recueil d’informations tant elles sont nombreuses. Effectivement, il paraît discutable de choisir telle méthode plutôt que telle autre. Cependant, comme pour certains indicateurs, certaines méthodes de collecte d’informations peuvent elles aussi faire partie de l’ensemble des « connaissances partagées » et s’imposer naturellement.
 - « l’examen de l’adéquation entre indicateurs et critères, ou la question du sens » : effectivement, cette question est cruciale. Il s’agit ici de juger si l’indicateur est bien représentatif de ce que nous nommons l’objectif (ou critère pour l’auteur). Par essence, une grande part de jugement est alors introduite ici. Notons que le jugement de la signification d’une information semble devoir se faire en même temps que le jugement de la représentativité de cette information par rapport à l’objectif, puisqu’il s’agit de retenir les informations (ou indicateurs) à associer au critère. Ce choix doit tenir compte à la fois de la signification, de la représentativité et de la disponibilité. Il s’agit bien ici d’une prise de position fondée sur jugement

d'une ou plusieurs personnes qui semble difficilement évitable.

Ajoutons à ces éléments de réflexion que l'information recueillie pour représenter un objectif peut être elle-même le résultat d'une autre évaluation, c'est-à-dire le résultat d'un jugement de valeurs amont. Prenons l'exemple de l'évaluation d'un élève afin de décider de l'attribution éventuelle d'un diplôme. Le jury va utiliser les informations, les indicateurs, que sont les notes obtenues dans les différentes matières. Certaines notes peuvent provenir d'un jugement (comme pour les matières littéraires ou artistiques) ayant auparavant été attribué par un enseignant de la matière (vu comme un expert). En ce sens, l'information utilisée par l'évaluant, le jury, est déjà le résultat d'un jugement de valeur. Cette situation est tout à fait acceptable si l'évaluant en est conscient, s'il sait qu'il a délégué une partie de sa responsabilité de jugement à un expert.

En conclusion, la subjectivité (au sens de Gerard) est présente dans l'évaluation. Nous insistons sur le fait que cette subjectivité est ici une série de jugements réalisés par une ou plusieurs personnes (à tour de rôle, par l'évaluant et l'expert). Elle n'est ni arbitraire ni péjorative. Elle est nécessaire à certaines étapes de l'évaluation. Cependant, il semble opportun que l'expert explique et justifie ses jugements notamment quand ceux-ci ne s'imposent pas d'eux même, ne résultent pas des « connaissances partagées ». Cela permet à l'évaluant de mieux maîtriser la qualité de l'évaluation. De son côté, si l'évaluant, de par son statut, n'a pas cette obligation de se justifier, nous estimons qu'il a parfois tout intérêt à mettre en avant, à définir précisément ses choix, ses prises de positions, ses jugements de valeurs car ceux-ci sont à l'origine même de la signification du résultat de l'évaluation. Cela contribue alors à la transparence du processus d'évaluation et donc des résultats qui en découlent. Evidemment, selon le contexte, cette transparence peut être souhaitée ou non.

1.3 Pourquoi évaluer ?

Quelles sont les motivations à la mise en place d'un processus d'évaluation ? A partir d'un ensemble d'objectifs prédéfinis, d'un référentiel, les buts de l'évaluation d'un système peuvent être de [9] :

- faire le constat, le diagnostic de l'état d'un système,
- communiquer avec transparence sur l'état d'un système,
- instaurer le dialogue et aider à la concertation,
- faciliter la négociation concernant les objectifs à retenir,
- sur une base commune d'objectifs, comparer un système avec un autre,
- réaliser le suivi de l'évolution d'un système au cours du temps,
- contrôler les dérives par rapport aux objectifs,
- identifier les faiblesses d'un système,
- alerter,
- révéler certains problèmes,
- faire apparaître de nouveaux enjeux décisionnels,
- prévoir l'évolution de l'état du système,
- anticiper les conséquences de certaines prises de décisions,
- comparer plusieurs scénarii.

Un vaste ensemble d'objectifs peut motiver l'évaluation d'un système. C'est tout ou partie de ces objectifs auxquels peut contribuer un processus d'évaluation. Il faut donc concevoir un modèle d'évaluation spécifique en fonction des objectifs visés.

1.4 Comment évaluer ?

1.4.1 Acteurs de l'évaluation

D'après Roy et Bouyssou [25], un acteur du processus de décision peut être défini comme « un individu ou un groupe d'individus qui, par son système de valeur, influence directement la décision, que ce soit au premier degré du fait de ses interventions ou au second degré par la manière dont il fait intervenir ceux d'autres individus ».

Par analogie, intéressons nous aux individus ou groupes d'individus influençant directement l'évaluation.

L'évaluant. Comme nous l'avons défini, l'évaluant est celui ou ce qui évalue. L'évaluant est celui ou ceux qui sont donc directement intéressés par le résultat de l'évaluation, ceux pour qui l'évaluation est nécessaire. Ils participent activement au processus d'évaluation (déclinaison de valeurs, consultation de résultats, etc) et sont donc des acteurs.

L'évalué. L'évalué est défini comme un système. Il fait l'objet de l'évaluation, ce sur quoi l'évaluation porte. Il semble d'ores et déjà important de bien définir les frontières de ce système. Cependant, il n'est pas rare que l'on cherche à évaluer les caractéristiques (aptitudes, connaissances, etc), d'une personne, d'un groupe de personnes (salariés d'une société, etc), ou plus généralement, d'un système dont les caractéristiques dépendent même partiellement d'individus. Dans ces cas, ces personnes sont acteurs à part entière du processus d'évaluation. En effet, leurs comportements, leurs passés, leurs productions, leurs dires, influencent l'état du système évalué. On doit alors s'attendre à une modification du comportement de l'évalué lorsqu'on construit et que l'on met en œuvre un tel modèle d'évaluation. Il s'agit d'effets rétroactifs pervers qu'il est tant que possible important d'identifier pour mieux anticiper leurs impacts.

Les experts. A plusieurs reprises, nous avons parlé de l'intervention d'experts. Ces personnes peuvent être amenées à participer au processus d'évaluation, notamment dans le cas où l'évaluant ne dispose pas de toutes les connaissances ou de toutes les compétences pour réaliser entièrement l'évaluation. Ces experts sont alors acteurs du processus d'évaluation.

Les hommes d'étude. Les hommes d'études sont des personnes qui vont tenter d'organiser le processus d'aide à l'évaluation. Ils vont être les interlocuteurs des différents acteurs entrant dans le processus. Ce sont eux par exemple qui vont tenter de présenter la méthodologie d'aide à l'évaluation et ses différentes étapes, de recueillir les valeurs des évaluateurs, d'instaurer un langage commun de communication entre les évaluateurs et les

experts. Nous pensons qu'au titre de ces participations, ils ne sont pas neutres et qu'ils doivent donc être considérés comme des acteurs à proprement dit du processus.

Les personnes extérieures. Certains résultats d'évaluations peuvent être diffusés a posteriori à des personnes complètement en dehors du processus d'évaluation. Par exemple, lorsqu'on rend publics les résultats du baccalauréat, tout le monde peut consulter ces résultats. On peut penser alors que ces personnes n'interviennent pas dans le processus d'évaluation puisque celui-ci est révolu au moment de la diffusion des résultats. Nous croyons cependant qu'elles peuvent influencer indirectement le processus d'évaluation si des acteurs (évaluant, évalué, experts) savent dès le début que les résultats seront diffusés. A ce titre, nous ne considérons pas ces personnes comme des acteurs à proprement dit, mais comme des acteurs indirects.

1.4.2 Processus d'aide à l'évaluation

Compte tenu de la façon dont nous avons défini l'évaluation, nous allons tenter de décrire le processus d'aide à l'évaluation (figure 1) grâce à plusieurs étapes que nous allons passer en revue dans cette section. Nous précisons les acteurs sollicités pour chacune d'elles. On peut d'ores et déjà remarquer que les hommes d'étude seront amenés à suivre chacune de ces étapes et sont donc à ce titre des acteurs tout au long du processus d'aide à l'évaluation.

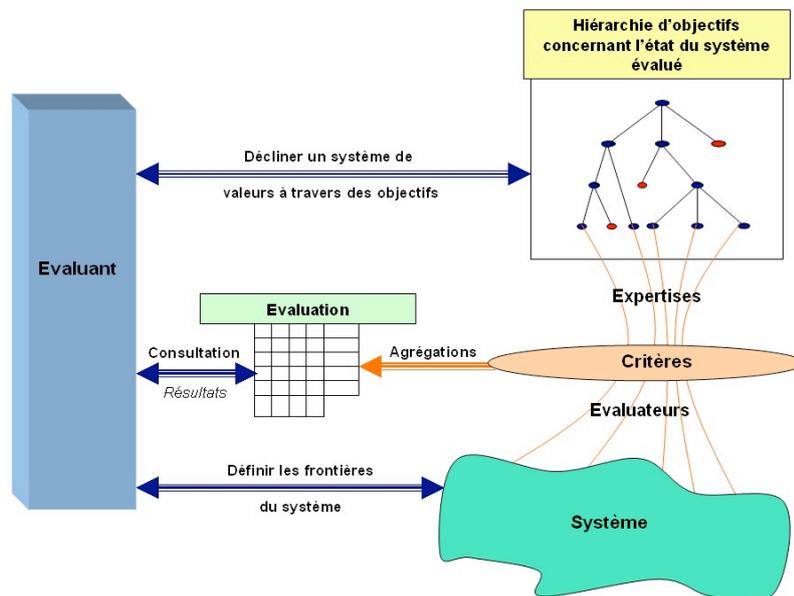


FIGURE 1 – Le processus d'aide à l'évaluation

Définir le système. Comme nous l'avons vu dans précédemment, il semble important de définir les frontières du système qui fait l'objet de l'évaluation. C'est donc assez naturellement que cette étape apparaît comme la première : avant toute chose, il est important d'identifier et de définir précisément ce que l'on cherche à évaluer. Cette étape sollicite l'évaluant. Il s'agit de discerner ce qui fait partie intégrante de ce que l'on cherche à évaluer, de ce qui ne le fait pas. Il s'agit aussi de définir qui sont les personnes que l'on inclut dans le processus, quel est l'horizon temporel de l'étude, etc.

Décliner un système de valeurs à travers des objectifs. La seconde étape nécessaire à l'aide à l'évaluation consiste à décliner un système de valeurs à travers un ensemble d'objectifs. Cette étape est cruciale car c'est elle qui va permettre de construire le référentiel qui va servir à l'évaluation. Elle fait bien entendu appel à l'évaluant. C'est en effet le système de valeurs de l'évaluant qui doit être décliné pour servir de référentiel. Les objectifs représentent les préoccupations de l'évaluant concernant l'état du système évalué. Notons que cette étape peut très bien être réalisée plusieurs fois au cours du processus. En effet, le système de valeurs de l'évaluant peut être amené à évoluer au cours du temps. Si des modifications sont apportées à l'ensemble des objectifs a posteriori, cela peut avoir des conséquences sur la suite du processus : il est donc parfois nécessaire de réitérer les étapes suivantes.

Sélectionner les critères mesurant l'atteinte des objectifs. Une fois les objectifs établis, il s'agit de sélectionner les critères permettant d'apprécier leur atteinte. Pour cela, l'évaluant peut choisir d'utiliser un ensemble d'informations faisant partie des « faits » comme introduit en section 1.2.3. Cependant, dans certains cas, l'évaluant doit faire appel à des experts qui sont plus à même de choisir les données à utiliser, les critères à retenir pour représenter les objectifs ou « valeurs ». Bien entendu, les experts ne doivent pas remettre en cause le bien fondé de l'objectif à représenter. S'il est difficile de trouver un critère parfaitement représentatif de l'objectif, l'expert peut proposer un ou plusieurs critères dont la signification se rapproche du mieux possible de celle de l'objectif. Dans ce cas, pour rendre la démarche opérationnelle, l'évaluant peut modifier certaines valeur à prendre en compte. L'évaluant revient alors dans sa phase de déclinaison du système de valeurs à travers des objectifs. Finalement, les deux acteurs de cette seconde phase sont l'évaluant et les experts, qui doivent travailler en interaction car l'évaluant doit faire un effort sémantique concernant la définition des objectifs retenus et l'expert doit lui tenter de s'approcher au mieux de ces définitions grâce aux critères.

Evaluer les performances sur les critères. Il s'agit ensuite d'appréhender la performance de l'évalué sur les différents critères. Cela peut se faire, par le biais de collecte d'information, grâce à des observations, des résultats de mesures physiques ou autre. Notons que c'est parfois l'évalué lui même qui produit le résultat servant à appréhender l'atteinte de l'objectif. Il est alors acteur dans cette étape. Mais dans la majorité des cas, c'est l'évaluant ou l'expert (selon le degré de connaissance de l'évaluant) qui tente de mesurer la performance de l'évalué sur un critère.

Consulter les résultats de l'évaluation. Enfin, la consultation des résultats est faite par l'évaluant ou a posteriori par des personnes extérieures au processus d'évaluation. Bien sur, certaines évaluations ne nécessitent pas d'agréger les données quand il n'y en a pas trop, ou quand il n'y pas trop d'objectifs. Cependant, il est souvent nécessaire de synthétiser l'information pour permettre une meilleure lecture des résultats.

1.5 Aide à la décision et aide à l'évaluation

Les principales méthodes d'agrégation multicritère ont été développées pour faire de l'aide à la décision. Dans ce cadre méthodologique, il s'agit entre autre de mettre en avant la ou les « meilleures décisions » à prendre parmi celles identifiées et cela par rapport aux préférences des décideurs. Mais notre problématique est différente. On ne se préoccupe pas des décisions possibles. Il s'agit d'évaluer une ou des situations environnementales relativement à un ensemble d'objectifs concernant l'état environnemental. Cela soulève certaines questions.

En premier lieu, nous présenterons l'aide à la décision de façon générale. Nous verrons comment on peut aider un décideur à appréhender un problème décisionnel, autrement dit, comment l'aide à la décision peut contribuer à éclairer un décideur. Pour cela, nous passerons en revue les différentes problématiques auxquelles les décideurs peuvent être confrontés.

Pour répondre à ces différentes problématiques, une étape nécessaire est d'évaluer les décisions envisagées. Dans la seconde section de cette partie, nous verrons en quoi consiste l'évaluation d'une décision en aide à la décision.

Enfin, nous verrons dans quelle mesure les problématiques de l'aide à la décision se rapprochent de la nôtre, visant à aider à évaluer.

1.5.1 Les problématiques de l'aide multicritère à la décision

Le but de ce paragraphe est de rappeler ce que signifie aider un acteur à prendre une décision. Pour cela, nous avons besoin d'introduire certains concepts de la méthodologie d'aide à la décision [25]. Nous ne reprendrons pas ici toutes les facettes de l'aide à la décision. Il existe, en effet, une vaste littérature concernant ce sujet et nous approfondirons les concepts dont nous avons besoin pour nos travaux tout au long de la section 3.3.

Le premier concept introduit ici est celui de l'action ou de l'alternative. Face à une situation donnée, un système, un décideur considère plusieurs options possibles, plusieurs décisions envisageables, qui auront chacune un impact sur l'état du système et modifieront l'ensemble de ses caractéristiques. L'ensemble des options possibles (aussi appelées *actions potentielles* ou *alternatives*) constitue l'ensemble A [25]. Ces *actions* peuvent être une à une exclusives, ou au contraire, compatibles.

Face à cela, l'aide à la décision peut être envisagée de plusieurs façons. La nature du problème, la « problématique », peut en effet varier. Si l'on est face au cas où le décideur doit choisir une action parmi toutes celles possibles (exemple : choix d'un tracé autoroutier), le problème n'est pas le même que si l'on peut choisir plusieurs actions (exemple :

définir un programme de recherche à partir d'un ensemble de projets possibles). Le décideur peut en effet s'intéresser uniquement à la ou les meilleures actions (problématique de la sélection) ou bien au contraire vouloir établir un ordre de toutes les actions pour identifier autant les meilleurs que les moins bonnes (problématique du rangement). Une troisième façon d'envisager d'aider un décideur est la classification. Cela consiste à classer les actions en les affectant chacune à une catégorie, une classe. Chaque classe a une signification par rapport au problème de décision (par exemple, pour un problème de choix de candidats on pourrait définir la catégorie des candidats irrecevables, celle des candidats admissibles, celle des candidats admis etc.). C'est la problématique de tri, que l'on appelle classement quand il n'existe pas de relation d'ordre entre les catégories. Enfin, le décideur peut choisir de se contenter d'une simple description de l'ensemble des actions, notamment quand le nombre des actions n'est pas trop élevé. C'est la quatrième problématique, celle de la description.

On distingue donc les quatre problématiques suivantes :

- de la sélection : mettre en avant la ou les meilleures actions pour aider à choisir l'action la plus intéressante aux yeux du décideur,
- de rangement : ranger toutes les actions de l'ensemble A selon un ordre de préférence (ou préordre),
- de tri : affecter chacune des actions à une des catégories prédéfinies,
- de la description : lister les performances des différentes actions selon chaque point de vue jugé pertinent.

Pour essayer répondre à ces quatre problématiques, il est nécessaire d'évaluer les différentes décisions possibles. Nous allons voir en quoi consiste ce type d'évaluation.

1.5.2 Evaluer une décision

Quelle que soit la problématique retenue, il s'agit d'essayer d'appréhender, pour chacune des actions les effets qu'elle va avoir sur l'état du système. En d'autres termes, c'est essayer de caractériser le changement d'état que va provoquer l'action. C'est-à-dire, la différence qu'il y aura entre l'état initial du système (ou état de référence) avec l'état final du système (après exécution de l'action). Bien sûr, on va restreindre l'étude de ces changements d'états à l'ensemble des points de vue qui importent aux yeux du décideur. On laissera de côté les conséquences qui ne l'intéressent pas. En ce sens, évaluer une action, c'est évaluer les changements d'états que provoque l'action et faisant partie des préoccupations du décideur. Remarquons que dans la plupart des cas, cela demande de réaliser un lourd travail de modélisation qui ne conduit pas toujours à des évaluations exemptes d'imprécision ou d'incertitude.

Introduisons maintenant le concept de critère pris dans le sens de Roy et Bouyssou [25]. Un critère est une échelle servant à représenter un point de vue qui importe aux yeux du décideur. Il permet la comparaison de deux actions selon le point de vue qu'il représente. Bien des précautions sont prises pour s'assurer de la cohérence de la famille de critères. Cette cohérence dépend du système de valeurs du décideurs. Ainsi, tout point de vue permettant de discerner deux actions au sens de la préférence du décideur doit être représenté par un critère (exhaustivité). L'ensemble des critères retenus doit

tant que possible vérifier d'autres bonnes propriétés (cohésion, non redondance, certaines formes d'indépendances).

En aide multicritère à la décision, le système de valeurs du décideur est appréhendé par le biais de son appréciation des conséquences des actions sur l'état du système. Il s'agit, pour les hommes chargés de l'étude de prendre en compte dans la modélisation tous les critères qui permettent au décideur de distinguer deux actions. Cette tâche aide à l'exploration d'une partie des valeurs du décideur : celles qui sont concernées lors de la comparaison de deux actions et de leurs conséquences. Ainsi, certaines valeurs peuvent à la fois être centrales dans un contexte décisionnel identifié et pour autant ne pas nécessairement apparaître dans le modèle d'aide à la décision.

1.5.3 Différences avec l'aide à l'évaluation

Notons avant tout, les travaux de Keeney [14] qui abordent les problèmes d'aide à la décision en s'intéressant en premier lieu aux objectifs. Nous trouvons cette approche très pertinente pour des raisons que nous exposerons dans la partie 2.

Dans le reste de la littérature, l'aide à la décision ne commence qu'assez rarement une étude en s'intéressant directement aux objectifs. En effet, elle s'intéresse tout d'abord aux actions [24]. Bien sûr, les critères sont ensuite choisis afin de refléter au mieux les préférences du décideur (restreintes au domaine d'étude) concernant l'évolution d'un état du système. Ils ne sont pas directement censés refléter les préférences concernant l'état du système à proprement dit, mais concernant un changement d'état. En aide à la décision, on n'explique pas directement la hiérarchie d'objectifs. De plus, ce qui est appréhendé pour construire les critères sont les performances des actions et non les caractéristiques du système. Notons aussi que la phase de consultation des résultats peut, selon la méthode d'agrégation utilisée et la forme du résultat auquel elle conduit, s'avérer assez complexe. On parle d'élaboration de recommandations. Il peut s'agir, en effet, d'exploiter une information assez difficile à interpréter directement.

Remarquons que l'existence des critères dépend forcément des objectifs du décideur, même si ceux-ci ne sont pas explicités dans une étude d'aide à la décision.

Dans un contexte décisionnel préalablement identifié, on peut concevoir qu'il existe une bijection entre la famille de critères au sens de l'aide à la décision et les objectifs concernant l'état du système au sens de l'aide à l'évaluation. Si l'on est capable de prévoir les conséquences des actions sur un état initial du système, on est aussi capable de prévoir un état final pour chacune des actions. Si l'on est capable de comparer entre eux chacun de ces états finaux par rapport aux objectifs du décideur, on est finalement capable de comparer indirectement entre elles toutes les actions. Fixons un état initial pour le système. Définissons un ensemble d'actions dont les conséquences sont connues (ou prévues). Associons à chaque action, un état final qui est le résultat de l'application de l'action sur l'état initial. Si la modélisation des critères et celle des objectifs sont parfaitement fidèles au système de valeurs du décideur (hypothèse forte), le résultat de la comparaison des états finaux par rapport aux objectifs doit être le même que celui de la comparaison des actions par rapport aux critères. Connaissant les objectifs d'un décideur, ne peut-on pas déduire une famille de critères reflétant les préférences de ce

décideur et permettant la comparaison des actions entre elles ?

Sans aller jusque là, on peut envisager de transposer les problématiques d'aide multicritère à la décision à notre contexte. Voyons alors à quoi correspondent les quatre problématiques présentées précédemment.

Pour la problématique de *sélection* il s'agira de faire ressortir le ou les meilleurs systèmes relativement aux objectifs du décideur. Ceci ne correspond pas parfaitement à ce que nous désirons mettre en œuvre ici. En effet, l'intérêt de cette problématique est de permettre d'y voir plus clair quand on a un grand nombre de systèmes, or nous nous intéressons surtout au cas où on a peu de systèmes à évaluer, voire un seul. D'autre part, si un système ne figure pas parmi les meilleurs de l'ensemble étudié, nous désirons tout de même savoir comment il se situe par rapport aux autres. Ce dernier point est rendu possible dans la problématique de *rangement*. Ainsi, cette seconde problématique se rapproche de nos préoccupations mais elle reste intéressante surtout dans le cas où l'on a un grand nombre de systèmes à comparer. Un inconvénient commun à ces deux problématiques est que l'évaluation du système se réalise relativement aux autres systèmes considérés dans l'étude. Il s'agit donc d'évaluations relatives. En conséquence, un système peut très bien être le meilleur d'un ensemble, tout en étant mauvais dans l'absolu.

Les deux autres problématiques semblent plus pertinentes pour répondre à nos ambitions. En effet, la problématique de *tri* qui consiste à affecter chaque système à une classe (ou catégorie) convient très bien à l'étude d'un seul système, ce qui est fondamental dans notre contexte. A cet effet, le choix du nombre de classes est doré et déjà identifié comme crucial. Si le nombre de classes est trop faible, il sera alors difficile d'évaluer avec finesse le ou les systèmes. Si à l'inverse, le nombre de classes est trop important, alors la signification de chacune d'elle pourra être remise en question. Malgré ces risques, un atout majeur de la problématique de tri est que l'on n'évalue pas les systèmes par rapport à d'autres systèmes, mais par rapport à des classes définies a priori, qui peuvent refléter les attentes du décideur concernant ses propres objectifs. Enfin, la problématique de la *description* permettrait de répondre à nos préoccupations en cela qu'elle informe sur l'état du système. En revanche, cette problématique ne synthétise pas l'information lorsqu'on est face à un ensemble d'objectifs trop important et trop hétérogène.

En définitive, la problématique de tri et la problématique de la description sont celles qui se rapprochent le plus de ce que nous désirons réaliser en aide à l'évaluation, appliquée dans notre cas à l'impact des transports sur l'environnement.

2 Structuration des objectifs

Une étape du processus d'aide à l'évaluation proposé dans la partie précédente consiste à modéliser un système de valeurs à travers des objectifs (cf. section 1.4.2). Dans cette seconde partie, nous souhaitons illustrer une méthode de structuration des objectifs en utilisant les travaux de Keeney [14] qui propose une manière d'aborder un problème décisionnel en réfléchissant tout d'abord sur les valeurs du décideur. Dans cette même perspective, notons que les travaux antérieurs de Gastaut et al [12] proposent une approche assez similaire dans un contexte de rationalisation des choix budgétaires.

Dans un premier temps (section 2.1), seront présentés certains concepts introduits par Keeney et les particularités de cette approche. Puis, une application de la structuration des objectifs concernant l'environnement et les transports sera réalisée grâce à une série d'entretiens exploratoires dans le but de pratiquer la méthode de structuration des objectifs et de vérifier sa faisabilité. Ces interviews seront présentées dans la section 2.2. En section 2.3 enfin, nous tenterons de montrer comment exploiter les résultats de ces interviews.

2.1 Approche par les valeurs

Dans la littérature sur les méthodes d'aide à la décision, nous allons nous intéresser particulièrement aux travaux de Keeney [14]. Dans « Value-focused thinking » l'auteur propose une approche des problèmes d'aide à la décision en se focalisant sur les valeurs du (ou des) décideur(s). Cette approche s'appuie sur une analyse structurée des objectifs reflétant le système de valeurs du décideur. Or, dans le paragraphe 1.4.2, nous avons décrit une étape qui consiste à traduire un système de valeurs à travers des objectifs. C'est en cela que la démarche de l'auteur nous intéresse. Dans cette partie, nous nous attacherons donc tout d'abord à présenter les principes sous-jacent à l'approche de Keeney. Nous verrons ensuite les différents types d'objectifs que Keeney distingue et les relations qui existent entre eux. Nous présenterons ainsi la méthode de structuration des objectifs proposée pour appréhender le système de valeurs du décideur.

2.1.1 Principes

Si l'on traduit « value-focused thinking » de manière littérale, cela donne « la pensée focalisée sur les valeurs ». Dans son ouvrage, Keeney oppose cela à une façon de penser bien plus classique en aide à la décision, qu'il appelle « alternative-focused thinking », autrement dit « la pensée focalisée sur les alternatives ». Il faut entendre ici, le mot « alternative », comme une action, une décision possible. Avant que le décideur ne commence à s'intéresser aux diverses alternatives qui s'offrent à lui, Keeney lui propose de faire tout d'abord le bilan de ses propres valeurs concernant le contexte décisionnel devant lequel il se trouve, en faisant abstraction des décisions possibles. Ce n'est que dans une étape ultérieure que les alternatives seront éventuellement analysées.

Il semble donc y avoir deux tendances extrêmes pour approcher un problème de décision : la première consistant à aborder le problème par les valeurs, l'autre par les

alternatives. Keeney (page 4)[14] fait une première différence entre ces deux approches : « Value-focused thinking essentially consists of two activities : first deciding what you want and then figuring out how to get it. In the more usual approach, which I refer to as alternative-focused thinking, you first figure out what alternatives are available and then choose the best of the lot ».

Dans la pratique, il s'agit souvent d'une utilisation conjointe de ces deux approches, à des degrés variables. Il est vrai que le plus souvent, en aide à la décision, le travail s'oriente dès le début vers l'analyse des alternatives. Même si celles-ci ne sont pas tout de suite évaluées, elles sont très vite recensées ou, à défaut, le type de décision (ou alternative) est lui très vite identifié. C'est généralement après ce travail que l'on s'intéresse aux valeurs du décideur. Il va, par exemple, être amené à décliner un ensemble de critères (financier, sociologique, géographique) permettant de comparer différents projets routiers donnés a priori. Bien entendu, le choix de ces critères peut très bien refléter avec finesse le système de valeurs du décideur. Mais si on ne s'intéresse pas aux valeurs du décideur avant l'analyse des alternatives, jamais dans l'étude le système de valeurs du décideur ne sera complètement explicité. Il sera bien sûr pris en compte tout au long du processus d'aide à la décision, notamment dans les choix de modélisation, mais ne sera pas nettement identifié. Il faudrait, pour cerner ce système de valeurs, savoir interpréter les différents choix de modélisation (comme le choix des critères et autres paramètres) ce qui est difficile.

Voyons maintenant comment l'auteur propose d'approcher les valeurs d'un décideur en tentant de structurer ses objectifs.

2.1.2 Les objectifs

Keeney (page 34)[14] définit un objectif somme suit : « An objective is the statement of something that one desires to achieve. It is characterized by two features : a decision context, an object and a direction of preferences. »

Selon lui, un objectif est donc quelque chose d'essentiel que l'on désire atteindre et qui se caractérise par :

- un contexte décisionnel,
- un objet de préoccupation,
- une direction de préférence.

Pour illustrer cela, prenons l'exemple du contexte décisionnel suivant : « la gestion du trafic routier ». Un objet de préoccupation peut alors être « les encombrements du trafic » pour lequel on peut avoir la direction de préférence : « diminuer ». Dans le contexte de gestion du trafic routier, un objectif est alors de diminuer les encombrements du trafic.

Le « contexte décisionnel » reste suffisamment large pour ne pas ni identifier a priori toutes les décisions possibles ni écarter d'emblée un ensemble de décisions, mais il précise cependant le champ d'action du décideur. En ce sens, on peut tout à fait rapprocher cette notion de « contexte décisionnel » avec notre notion de « contexte » en évaluation qui signifie, l'activité plus large dans laquelle s'insère le processus d'évaluation.

Keeney ajoute que pour rechercher les objectifs, il n'est utile ni d'identifier la mesure permettant de représenter l'objectif, ni de préciser les différents niveaux ou paliers qui

serviront de références sur cette mesure pour indiquer l'atteinte de l'objectif. Ce point est important car il facilite l'étape de définition des objectifs : on ne se préoccupe encore ni de la façon dont on va construire le ou les critères que l'on va retenir pour appréhender le niveau d'atteinte de ces objectifs.

Voyons maintenant la façon dont Keeney aborde le problème de structuration des objectifs. Il commence par distinguer deux types d'objectifs. Les objectifs de fin (traduction de « fundamentals objectives », que l'on pourrait aussi traduire par objectif de but) et les objectifs de moyens (« means objectives »).

2.1.3 Les objectifs de fin et de moyen

« The fundamental objective characterizes an essential reason for interest in the decision situation. » Keeney (page 34)[14]

Pour l'auteur, quand l'objet représente une raison essentielle qui fait que le décideur s'intéresse à la situation décisionnelle, il s'agit alors d'un objectif de fin. Un objectif de fin est décomposable. En effet, « limiter l'encombrement du trafic routier » peut se décomposer par exemple en deux sous-objectifs : « limiter l'encombrement du trafic routier dans les villes » et « limiter l'encombrement du trafic routier en dehors des villes ». Ces deux objectifs sont deux facettes de l'objectif plus général « limiter l'encombrement du trafic routier ». Ils sont aussi objectifs de fin mais d'un niveau plus bas. Ces nouveaux objectifs de fin de plus bas niveau peuvent à leur tour être décomposés en objectifs de fin de plus bas niveau, jusqu'au moment où ils ne seront plus décomposables. À l'inverse, plusieurs objectifs de fin peuvent être regroupés au sein d'un même objectif s'ils sont diverses facettes d'un même objet plus général.

Nous avons choisi le mot « fin » pour traduire « fundamental » car ces objectifs représentent une fin en soi pour le décideur. En opposition à ce concept, Keeney introduit l'objectif de moyen, traduction de « means objective », qui ne représente pas une fin en soi pour le décideur, mais un moyen d'atteindre un objectif de fin : « A means objective is of interest in the decision context because of its implications for the degree to which another (more fundamental) objective can be achieved. Simply stated, the means objectives are important because they are means to the achievement of the fundamental objectives » Keeney (pages 34-35)[14].

Dans un certain contexte, l'objectif visant à limiter les encombrements du trafic routier est une fin en soi et donc un objectif de fin. Imaginons maintenant l'objectif consistant à améliorer la synchronisation des feux dans la ville. Cet objectif contribue à limiter les encombrements du trafic routier. Il ne s'agit pas d'une fin en soi. C'est un objectif de moyen. Il existe bien un lien de cause à effet entre la synchronisation des feux (l'objet de l'objectif de moyen) et l'encombrement du trafic routier (l'objet de l'objectif de fin).

Keeney précise aussi qu'un objectif de moyen peut être lié à un ou plus objectif de fin. Il se peut parfois qu'un moyen mis en œuvre pour l'atteinte d'un objectif de fin ait en effet des conséquences concernant un autre aspect que l'objectif de fin en question. Il peut donc influencer à ce titre sur différents objectifs de fin, positivement (dans la direction de préférence) comme négativement (dans la direction inverse de la préférence). Améliorer la

synchronisation des feux sur un réseau routier influence positivement l'objectif « limiter l'encombrement du trafic routier ». Imaginons maintenant que le décideur intéressé par le contexte décisionnel du trafic routier a un objectif visant à « limiter les dépenses ». Cet objectif peut être un objectif de fin. Or améliorer la synchronisation des feux sur un réseau routier augmente les dépenses, ne serait-ce que ponctuellement (étude, matériel, etc). Cet objectif de moyen influence donc (et cette fois négativement) l'objectif « limiter les dépenses ». Autrement dit, un objectif de moyen peut avoir des répercussions sur le niveau d'atteinte de plusieurs objectifs de fin. Si l'on tente de représenter ces diverses relations on obtient la figure 2. La schéma figure 3 est une représentation plus générale de ces relations.

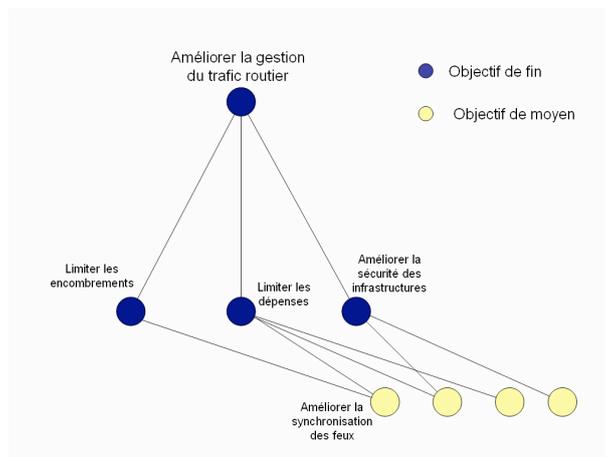


FIGURE 2 – Relations entre objectifs de fins et de moyens : gestion du trafic routier

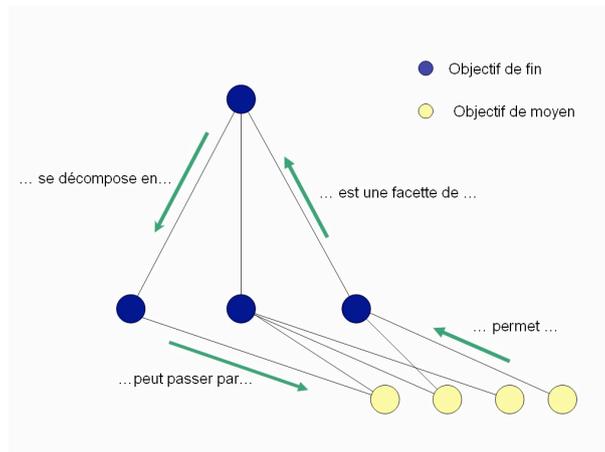


FIGURE 3 – Relations entre objectifs de fins et de moyens : schéma général

2.1.4 Objectif stratégique

Keeney définit l'objectif stratégique comme un objectif de fin qui a la particularité d'être invariant au cours du temps et commun à plusieurs situations décisionnelles. Ils sont souvent des objectifs de fins de plus haut niveau. Pour une entreprise, il s'agit par exemple, d'être rentable et d'être pérenne. Pour une personne, cela peut être de vivre longtemps, dans de bonnes conditions et en bonne santé. Concernant un décideur, ce type d'objectifs semble donc un peu trop personnel et de ce fait certainement trop difficile à révéler pour faire l'objet de ce travail (même si cela peut s'avérer très intéressant).

Nous ne préconisons donc pas l'exploration d'objectifs stratégiques dans notre cas. Le plus haut niveau d'objectif de fin à explorer se doit de concerner la gestion de l'environnement et des transports, qui est notre thème d'application.

2.1.5 Révéler les objectifs

Plusieurs démarches sont proposées par Keeney [14] pour structurer les objectifs d'un décideur. Ces méthodes sont là pour alimenter la réflexion puis le discours du décideur. Parallèlement à ce travail, la personne chargée de mener à bien l'étude se doit de noter tous les objectifs qu'elle identifie, en distinguant les objectifs de fin de ceux de moyen. Cette étape est nécessaire pour débiter la suivante qui consiste à structurer et à explorer l'ensemble de la hiérarchie des objectifs du décideur.

Selon l'auteur, un premier moyen d'identifier des objectifs est d'entamer une discussion à propos de la situation décisionnelle. Cela permet souvent de faire apparaître le contexte décisionnel dans lequel on se place ainsi que certains objectifs. On peut ensuite simplement demander au décideur les objectifs qu'il souhaite atteindre dans ce contexte, sans pour autant lui demander l'importance relative qu'il apporte à chacun d'eux. On peut aussi lui faire considérer un ensemble d'alternatives tant fictives que réelles. On peut alors lui demander : « quels peuvent être les objectifs de telle alternative ? » ou « qu'est-ce qui permet de distinguer telle alternative de telle autre ? » ou encore « qu'est-ce pour vous une alternative parfaite, la pire alternative ? ».

Remarquons que l'*approche par les valeurs* que propose Keeney n'hésite pas à s'appuyer sur une *approche par les alternatives* pour révéler les objectifs. En cela, ces deux approches extrêmes (introduites en section 2.1.1) ne sont pas incompatibles.

Pour révéler les objectifs, l'auteur propose d'explorer les problèmes auxquels est confronté le décideur. Une autre façon peut être d'observer les conséquences des alternatives, c'est-à-dire leurs différents impacts. Il faut alors demander au décideur si ces conséquences importent ou non à ses yeux. On peut aussi explorer les buts ou contraintes. Selon Keeney, les buts diffèrent des objectifs dans le sens où les buts sont des objectifs avec un niveau souhaité d'atteinte. Les contraintes sont aussi des objectifs mais avec un niveau d'atteinte requis. En décelant les buts ou les contraintes, on identifie donc des objectifs. On peut aussi demander au décideur de s'exprimer en se mettant à la place d'une autre personne concernée par la décision, afin de changer de perspectives, de vision des choses. De plus, dans l'étape de construction de la hiérarchie (étape décrite dans la section suivante) il est fréquent qu'apparaissent de nouveaux objectifs jusque là non

identifiés. Enfin, dans l'étape qui consiste à sélectionner les critères mesurant l'atteinte d'un objectif (étape décrite dans la section 1.4.2), il est fréquent pour un même objectif de disposer de différentes échelles ne mesurant pas tout à fait la même chose. Quand c'est le cas, c'est qu'il existe en fait plusieurs objectifs souvent très proches l'un de l'autre. Ils sont ainsi révélés et il faut alors les spécifier.

2.1.6 Construire l'arbre d'objectifs

On peut rattacher chacun des objectifs de moyen cités à au moins un objectif de fin en posant la question « Pourquoi cet objectif (est-il important) ? ». De même, à partir d'un objectif de fin, on peut construire sa hiérarchie supérieure (Bottom-Up) en demandant « Pourquoi cet objectif ? ». Pour explorer en profondeur l'arbre des objectifs de fin (Top-Down), on demandera « En quoi cet objectif est-il important ? », où bien « Quelles facettes de cet objectif sont importantes ? ». Enfin, pour faire apparaître des objectifs de moyens supplémentaires, on demandera, à partir d'un objectif de fin « Comment atteindre un tel objectif ? » Ces questions sont représentées dans la figure 4.

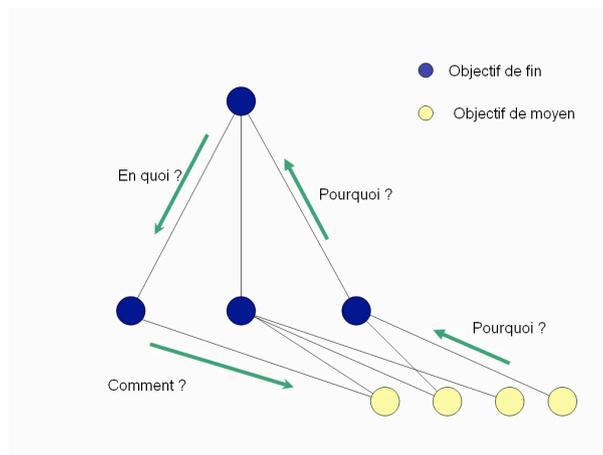


FIGURE 4 – Questions permettant de parcourir une structure d'objectifs

2.1.7 Résultats : l'arbre des objectifs de fin

De la définition des liens entre objectifs faite par Keeney, il résulte que l'ensemble des objectifs de fin constitue une hiérarchie, un « arbre » au sens de la théorie des graphes.

Rappels concernant la théorie des graphes : arbre. Les concepts présentés ici proviennent de la théorie des graphes (Berge [3]). Un graphe permet de décrire un ensemble d'objets et leurs relations, c'est-à-dire les liens entre les objets. Un graphe G est un couple (V, E) où :

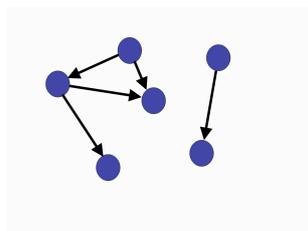


FIGURE 5 – Graphe orienté, non connexe

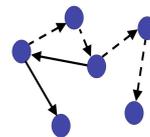


FIGURE 6 – Chemin simple dans un graphe orienté connexe

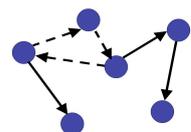


FIGURE 7 – Circuit dans un graphe orienté connexe

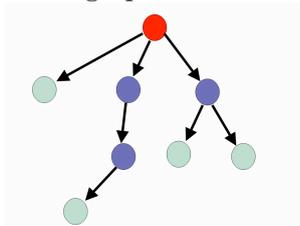


FIGURE 8 – Arbre

- V est un ensemble (fini) d'objets. Les éléments de V sont appelés les sommets du graphe.
- E est sous-ensemble du produit cartésien $V \times V$. Les éléments de E sont appelés les arêtes du graphe.

Un « chemin » est une suite de sommets reliés par des arêtes. Un graphe est dit « connexe » dès lors qu'il existe au moins un chemin permettant de relier tout couple de sommets. Le graphe de la figure 5 n'est pas connexe tandis que ceux des figures 6 et 7 le sont. Un chemin p est simple si chaque arc du chemin est emprunté une seule fois, comme le chemin représenté en pointillés sur la figure 6. Un cycle est un chemin simple finissant à son point de départ. La figure 7 fait apparaître un cycle en pointillés.

Un arc est une arête dotée d'un sens de parcours. Un graphe constitué de sommets et d'arcs est dit « orienté ».

Un arc e du graphe G est un couple (u, v) de sommets. Les sommets u et v sont les extrémités de l'arc dont l'orientation est alors de u vers v . On dit que u et v sont des sommets voisins. La figure 5 est un graphe orienté pour lequel chaque arc est représenté par une flèche décrivant le sens de parcours, l'orientation.

Le degré entrant d'un sommet v est le nombre d'arcs dont l'extrémité finale est v .

Un arbre est un graphe orienté connexe sans circuit où chaque sommet est de degré entrant au plus 1 (figure 8). Sous ces conditions, un seul sommet est de degré entrant 0 : c'est la racine, notée r , sommet sans prédécesseur. Pour un arbre T de racine r :

- le père d'un sommet e est l'unique voisin de e sur le chemin de la racine à e . La racine r est le seul sommet sans père,
- les fils d'un sommet e sont les voisins de e autres que son père,
- une feuille est un sommet sans fils.

Résultats. Selon les quelques définitions venant d'être rappelées, le résultat de la structuration des objectifs de fin est un arbre dont chaque sommet représente un objectif de fin. Chaque arc représente le lien de décomposition allant de l'objectif décomposé vers celui qui le décompose. Chaque feuille est un objectif de fin non décomposable. On parlera par la suite, dans un arbre d'objectifs :

- O' descendant de l'objectif O : tel qu'il existe un chemin allant de O vers O' ,
- O objectif décomposable : O n'est pas une feuille (O a au moins un fils).

Bien entendu, l'ensemble des objectifs de moyens et leurs liens (entre eux ou avec les objectifs de fin) ne composent pas un arbre, notamment car les cycles sont permis.

2.2 Interviews

Cette section concerne une application de la méthode de structuration des objectifs proposée par Keeney dans le domaine de l'environnement et des transports. Pour cela, une série de sept interviews a été réalisée.

Nous établirons tout d'abord la consigne de départ pour les interviews afin de cadrer le discours des personnes interrogées. Nous expliciterons ensuite le protocole d'interview. Enfin, nous présenterons les résultats obtenus pour tenter de construire une structure d'objectifs de synthèse.

Le but de ce travail exploratoire est de valider le mode de construction de la structure d'objectifs de fins et de tester son applicabilité au domaine de l'environnement et des transports. Nous ne prétendons donc pas être exhaustif quant aux objectifs environnementaux.

2.2.1 Interviewés

Le projet PIE vise à construire un outil permettant l'évaluation de l'impact des transports sur l'environnement. Les demandeurs de ce type d'outils sont principalement des responsables politiques du domaine de l'environnement et/ou des transports, mais comme nous l'avons vu, ils ne sont pas les seuls utilisateurs potentiels. Les représentants des acteurs économiques et sociaux sont eux aussi susceptibles d'utiliser l'outil. Qui que soit l'utilisateur de l'outil, celui-ci sera amené dans une première étape à décliner ses valeurs à travers un ensemble d'objectifs mis à sa disposition. Il faut donc que cet ensemble d'objectifs soit le plus exhaustif possible pour pouvoir être adaptable au plus grand nombre de système de valeurs d'utilisateurs.

Pour construire un ensemble d'objectifs qui soit le plus exhaustif possible, il faudrait interviewer des échantillons d'individus représentatifs de chacune des catégories d'acteurs concernés par l'environnement et les transports (responsables politiques, acteurs économiques, acteurs sociaux). Nous n'avons pas orienté notre travail vers cette perspective car notre but est avant tout une validation méthodologique. Ainsi, nous ne recherchons pas à cette étape l'exhaustivité des valeurs issues des interviews, mais simplement, des discours les plus vastes et différents possibles. Nous avons pour cela choisi d'interviewer des responsables publics ou élus. Le critère de sélection est qu'ils aient, de près ou de loin, des préoccupations et responsabilités dans le domaine des transports

et/ou de l'environnement. Nous avons parfois rencontré leurs conseillers techniques qui sont souvent plus disponibles pour ce genre d'entrevues. Ils se sont dans ce cas exprimés au nom du décideur pour qui ils travaillent.

Les personnes rencontrées étaient chargées d'affaires dans des villes de tailles différentes (arrondissement parisien, ville de Lyon, ville de région parisienne, ville moyenne de province, petite commune de province). Les tendances politiques ne sont pas toutes représentées ; nous n'avons pas choisi les interviewés en fonction de ce critère, mais selon leur disponibilité et leur fonction dans la ville.

Bien sûr, nous n'avons pas non plus été exhaustifs concernant le type de responsable. Nous n'avons par exemple pas rencontré de personnes en charge d'une région et encore moins de parlementaires. Nous nous sommes contentés de responsables au niveau de la ville pour réaliser cette validation méthodologique.

2.2.2 Consigne de départ

Les valeurs que nous cherchons à explorer durant chaque interview sont restreintes à celles concernant l'impact des transports sur l'environnement. Il est vrai que d'un point de vue pratique, et compte tenu de la disponibilité des personnes rencontrées, il n'aurait pas été concevable de leur demander de nous consacrer le temps nécessaire pour parcourir l'ensemble du champ « transport » qui intègre tant les aspects environnementaux que les aspects sociaux et économiques. Nous avons donc finalement retenu la consigne suivante : « Quels sont, pour vous, les objectifs environnementaux à prendre en compte dans le suivi d'une politique de transport ? »

Nous explorons ainsi des objectifs concernant l'environnement et non les objectifs concernant les transports. Mais de cette façon, nous restreignons notre exploration aux aspects de l'environnement sur lesquels les transports ont un impact. A notre sens, cet ensemble d'objectifs convient bien au domaine de notre étude car il représente un référentiel qui devrait permettre d'évaluer l'impact des transports sur l'environnement.

2.2.3 Protocole

Avant de rencontrer les personnes que nous souhaitons interviewer, nous avons réalisé une série de trois interviews test, dans le but de s'entraîner, d'anticiper d'éventuelles difficultés, de fixer le temps nécessaire au bon déroulement du travail de structure des objectifs. Les gens rencontrés dans ce cadre n'avaient pas de responsabilités publiques. Ils étaient cependant bien informés des préoccupations environnementales liées aux transports. Cela nous a permis de bien cadrer le protocole d'entretien exploratoire. Ainsi, le protocole de l'interview consiste en deux choses : dans un premier temps la préparation de l'interview puis l'interview en lui-même.

2.2.4 Préparation des interviews

Nous avons contacté les personnes rencontrées une première fois par courrier électronique en leur joignant une note explicative concernant l'entretien qui présentait :

Les objectifs. « Dans le cadre du projet de recherche PIE (Prospectives et Indicateurs des impacts des transports sur l'Environnement), je⁶ réalise une thèse à l'Inrets concernant la conception d'un outil informatique. Celui-ci est destiné à être utilisé par des décideurs (élus) et servira au suivi d'une politique de transports grâce à l'évaluation de situations environnementales par rapport à des objectifs. Afin de préparer au mieux les fonctionnalités de l'outil, nous souhaitons explorer les différents objectifs qu'il serait possible d'avoir pour un tel suivi. Pour cela, nous avons choisi de rencontrer directement les personnes en charge de prendre des décisions dans ce domaine. Nous n'attendons aucunement un discours axé sur les connaissances des phénomènes environnementaux et des transports que pourrait avoir un technicien, mais voulons tenter d'appréhender les jugements de valeurs (différents objectifs, priorités, etc) d'élus. Ainsi, la consigne de départ est : quels sont pour vous les objectifs environnementaux à prendre en compte dans le suivi d'une politique des transports ? »

La confidentialité. « Il s'agit d'un travail de recherche exploratoire. A ce titre, l'utilisation des résultats de ces interviews se fera de façon anonyme et l'identité des interviewés restera confidentielle. »

L'organisation. « La rencontre n'excèdera pas une heure. A l'issue de la rencontre, une version de la « structure des objectifs » sera établie. Celle-ci sera transmise a posteriori à l'interviewé. Selon son désir des modifications pourront être apportées, où, le cas échéant, une nouvelle rencontre pourra être organisée. »

Le résultat. « Le résultat final de la recherche sera une synthèse générale de toutes les interviews. Nous tenterons de construire un arbre générique des objectifs, qui, mise au sein de l'outil, devrait permettre à l'utilisateur final de décliner ses propres objectifs environnementaux. »

2.2.5 Technique d'interviews

L'entretien en lui-même se découpe en deux phases.

La première phase sert à faire apparaître un premier ensemble d'objectifs. Elle consiste à laisser parler l'interviewé en faisant des relances non directives. Une prise de notes permet alors de faire un premier recensement des objectifs qui apparaissent dans le discours de l'interviewé. Pour cela, la consigne est tout d'abord donnée à l'interviewé : « Quels sont pour vous les objectifs environnementaux à prendre en compte dans le suivi d'une politique des transports ? ». On distinguera doré et déjà, parmi les réponses, les objectifs de fins et de moyens. Dans cette phase, les interventions de l'intervieweur ne seront là que pour demander l'explication sémantique des termes employés afin d'assurer une bonne compréhension. En cas d'arrêt ou de blocage, l'intervieweur pourra utiliser des relances de type : « N'y a-t-il pas d'autres aspects environnementaux qui dépendent

6. Benjamin Rousval

des transports ? Si oui, sont-ils pour vous des objectifs à prendre en compte ? ». Cette phase a duré entre 15 minutes et 35 minutes selon les personnes rencontrées.

La seconde phase de l'entretien consiste à explorer la structure d'objectifs de l'interviewé à partir des éléments cités dans la première phase. Ainsi, on peut rattacher un objectif de moyen cité à un objectif de fin en posant la question « Pourquoi cet objectif ? ». De même, à partir d'un objectif de fin cette fois, on peut construire son supérieure (Bottom-Up) en demandant « Pourquoi cet objectif ? ». Pour explorer en profondeur l'arbre des objectifs de fin (Top-Down), on demandera « En quoi cet objectif est-il important ? », ou bien « Quelles facettes de cet objectif sont importantes ? ». Ceci est représenté dans la figure 4 à l'exception de la question « Comment ? » car nous avons choisi de ne pas explorer les objectifs de moyen. Une première raison à ce choix est que les personnes interviewées ne sont pas censées être des experts des transports et de l'environnement, mais des personnes devant exprimer un jugement de valeur. Or, nous avons vu dans la section 2.1.3 qu'il existe un lien de type causal entre un objectif de moyen et l'objectif de fin auquel il se rattache. La relation « cause à effet » est un phénomène principalement connu par les experts, faisant donc partie des connaissances scientifiques. Pour explorer cela, il s'agit tout autant de réaliser une recherche bibliographique de rapports d'expertises concernant l'efficacité de moyens connus, des mesures qui ont déjà été prises, qui sont à l'étude, qui ont été proposées, que d'interviewer des décideurs. Pour une raison pratique ensuite, il semblait difficile d'envisager de tout faire bien en une heure de temps qui est la durée que l'on fixe pour chaque interview.

Autrement dit, nous explorons avant tout les « valeurs » des interviewés et non leur degré de connaissance des phénomènes environnementaux liés aux transports.

2.3 Résultats

2.3.1 Les différents arbres

Voici les fonctions des sept personnes interviewées (entre mai et septembre 2003) :

- Interview 1 : Maire d'une petite ville de province
- Interview 2 : Adjoint au maire dans une petite ville
- Interview 3 : Maire d'une ville en région parisienne
- Interview 4 : Elu dans un arrondissement parisien
- Interview 5 : Conseillé à la mairie d'une grande ville
- Interview 6 : Conseillé à la mairie d'une autre grande ville
- Interview 7 : Personne dans l'opposition à la mairie d'un arrondissement parisien.

Concernant les durées effectives des rencontres, seuls les interviews 3 et 4 ont excédé une heure. En revanche, l'interview 5 a été d'une durée plus courte (aux alentours de 45 minutes).

Ils se sont déroulés en privé dans des bureaux sauf l'interview 6 qui a eu lieu dans un lieu public.

Nous ne présenterons pas ici les structures d'objectifs obtenues lors de tous les entretiens. Nous avons choisi d'en présenter deux à titre d'exemples (figures 9 et 10). Les arbres présentés concernent uniquement les objectifs de fin. Par exemple, la figure 9 se lit

de la manière suivante : l'objectif « Maîtriser l'environnement » passe par les objectifs « à l'échelle planétaire » et « à l'échelle locale ». Ce dernier a lui même deux sous-objectifs : « concernant le milieu naturel » et « concernant la société humaine ».

Le but du travail est maintenant d'utiliser les résultats des différentes interviews, et tenter de les synthétiser en un seul arbre d'objectifs de fin.

Nous allons tout d'abord présenter la méthode que Keeney propose pour synthétiser plusieurs arbres d'objectifs en un seul, et discuter de l'applicabilité de cette approche dans notre cas. Dans un second temps, nous tenterons de réaliser la synthèse proprement dite. Enfin, nous présenterons l'arbre d'objectifs obtenu après réunion des sept arbres.

2.3.2 Méthode de synthèse des arbres

Pour réunir plusieurs arbres d'objectifs en un seul, Keeney [14] propose la définition d'une opération d'union. Nous allons discuter de l'applicabilité de cette union à notre cas.

L'union. L'opération d'union que propose Keeney [14] permet de réunir deux arbres d'objectifs de fin. Pour cela, il considère les deux arbres ayant été bâtis dans un même contexte de décision en s'intéressant tout particulièrement aux objectifs de plus haut niveau. Cela forme deux ensembles distincts. Il réalise alors l'union de ces deux ensembles, en prenant la précaution de ne pas répéter deux objectifs suffisamment proches sémantiquement l'un de l'autre (ou synonymes). Il choisit ensuite un objectif de l'ensemble ainsi obtenu. Il fait alors l'union de tous les sous-objectifs directement rattachés à cet objectif tant dans le premier arbre que dans le second. Il répète enfin cette opération pour tous les objectifs du premier ensemble obtenu, et obtient ainsi tout les ensembles d'objectifs de second niveau. Il réalise à nouveau cette tâche pour le troisième niveau et ainsi de suite.

Applicabilité. Nous pouvons tenter d'appliquer cette méthode à notre cas. Ici l'union que nous devons réaliser se fera sur sept arbres d'objectifs. Bien sûr, la consigne initiale a bien été la même pour chacun des interviewés et cela permet de considérer le contexte décisionnel comme identique pour chaque arbre. Notons cependant une particularité sur le résultat des travaux. En effet, nous constatons l'apparition de « séparateurs ». Ce que nous appelons séparateurs, sont des aspects permettant de décomposer un objectif en plusieurs sous-objectifs. Par exemple, pour l'objectif initial « maîtriser l'environnement », on peut proposer comme séparateur, l'aspect temporel et obtenir ainsi trois sous-objectifs comme : « maîtriser l'environnement à court terme », « maîtriser l'environnement à moyen terme » et « maîtriser l'environnement à long terme ». Les séparateurs qui apparaissent le plus souvent dans le résultat des nos entretiens sont l'échelle d'observation temporelle (long terme, court terme, etc) et l'échelle d'observation géographique (niveau planétaire, niveau national, etc). Or, utiliser ces séparateurs dans un ordre, ne produit pas le même arbre que si on les utilise dans un autre ordre, même si d'un point

Maîtriser l'environnement

- A l'échelle planétaire
 - Préserver l'avenir
 - Pour les ressources naturelles
 - Sols
 - Eaux
 - Espèces
 - Energies
 - En terme de consommation
 - En terme de production de déchets
 - Pour le patrimoine humain
 - Concernant le patrimoine historique
 - Concernant le patrimoine culturel
 - Améliorer la situation actuelle concernant
 - Réchauffement planétaire
 - Dérèglement climatique
 - Couche d'ozone
- A l'échelle locale
 - Concernant le milieu naturel
 - Les végétaux
 - Protéger
 - Préserver
 - Les animaux
 - Protéger
 - Préserver
 - Le paysage
 - Concernant la société humaine
 - Protéger les bâtis
 - Habitats
 - Patrimoine
 - Limiter les effets sur la santé
 - Pour les population à risque
 - Pour les zones à forte densité
 - Limiter la gêne
 - Due aux bruits
 - Nocturnes
 - Pics

FIGURE 9 – Résultat de l'interview 1

Maîtriser l'environnement

Améliorer l'environnement d'aujourd'hui

Améliorer la santé publique

La santé routière

Pour les usagers

Pour les non usagers

Les effets de la pollution sur la santé

Pour les populations à risque

Pour le reste de la population

Améliorer la qualité de la vie urbaine

En limitant la gêne

Liée au bruit

Liée aux odeurs

Liée aux fumées

En facilitant les transports

En commun

Individuels

En conservant les bâtis

Améliorer la situation planétaire

Du point de vue des pollutions maritimes

Accidentelles

Frauduleuses

Du point de vue des déchets

Du point de vue des consommations d'énergies

Du point de vue du réchauffement planétaire

Préserver l'environnement futur

Préserver le patrimoine

Ses monuments

Ses espaces naturels

Ses paysages

Eviter les conséquences irréversibles

Disparition d'espèces naturelles

Animales

Végétales

FIGURE 10 – Résultat de l'interview 2

de vue sémantique rien n'est fondamentalement différent. Voici les deux possibilités de notre exemple.

1er cas :

Niveau planétaire
 Court terme
 Long terme
Niveau national
 Court terme
 Long terme

2ème cas :

Court terme
 Niveau planétaire
 Niveau national
Long terme
 Niveau planétaire
 Niveau national

Ici, le problème de l'union est que des arbres ont été construits avec tout ou partie des séparateurs utilisés dans l'ordre qu'a choisi l'interviewé. Donc, certains des sous-objectifs issus de l'utilisation de ces séparateurs interviennent à des niveaux potentiellement différents. Il y a donc un risque non négligeable de répétitions lors des opérations d'unions telles que définies par Keeney qui s'effectuent aux différents niveaux de l'arbre. Pour pallier cet effet que nous nommons « libre choix de l'ordre des séparateurs », nous pensons qu'il est possible de normaliser les arbres obtenus en les modifiant, sans changer le reste de l'arbre. Ainsi, nous proposons de transformer l'arbre du premier interview qui débutait comme suit :

A l'échelle planétaire
 Préserver l'avenir
 Améliorer la situation actuelle
A l'échelle locale
 Pour la situation actuelle

En inversant les séparateurs Échelle géographique / Échelle de temps nous obtenons alors :

Améliorer la situation actuelle
 A l'échelle planétaire
 A l'échelle locale
Préserver l'avenir

A l'échelle planétaire

Nous obtenons grâce à cette transformation un arbre qui se rapproche un peu plus de l'arbre de notre second interviewé, sans pour autant avoir changé les objectifs, c'est-à-dire les valeurs de la personne en question. Cependant, si l'on regarde l'arbre, dans son volet « Améliorer l'environnement d'aujourd'hui » il distingue « Améliorer la santé publique », « Améliorer la qualité de la vie urbaine » puis « Améliorer la situation planétaire ».

Améliorer l'environnement d'aujourd'hui
Améliorer la santé publique
Améliorer la qualité de la vie urbaine
Améliorer la situation planétaire

Quand on entre dans le détail des sous objectifs des deux premiers volets, on s'aperçoit vite qu'il s'agit d'objectifs de santé « locale », et de qualité de vie « locale » (niveau urbain). Cela, même si le niveau « local » n'est pas précisé. On peut transformer l'arbre sans en modifier le sens comme suit :

Améliorer l'environnement d'aujourd'hui
A l'échelle locale
Améliorer la santé publique
Améliorer la qualité de la vie urbaine
A l'échelle planétaire

On obtient ainsi les mêmes séparateurs que dans le premier arbre. L'opération d'union se fera d'autant plus facilement. Ces deux opérations de normalisation seront donc intéressantes dans notre cas. Nous y procéderons pour chacun des arbres, avant de réaliser à l'union. Plus précidément, il s'agit tout d'abord de normaliser des deux premiers arbres, puis de les unir. Ensuite, nous normalisons le troisième arbre avec celui obtenu par l'union des deux premiers. Une fois normalisés, nous les unissons et ainsi de suite jusqu'au septième arbre.

2.3.3 Application

La figure 11 présente le résultat de l'union des sept arbres d'objectifs de fins. Celui proposé ici est limité à cinq niveaux de profondeurs, par souci de clarté. Dans sa version complète ([22]), l'arbre comporte 92 feuilles.

Maîtriser l'environnement

A l'échelle planétaire

Conserver un environnement favorable à la vie humaine

 Limiter l'effet de serre

 Limiter le dérèglement climatique

 Protéger la couche d'ozone

 Eviter la prolifération de grandes pandémies

 Eviter le pillage des peuples

 Eviter les guerres pétrolières

Conserver les ressources naturelles

 Limiter l'extinction d'espèces naturelles

 Limiter l'extinction de milieux naturels

Limiter les consommations d'énergies

Limiter la pollution maritime

Limiter la production de déchets persistants

Limiter la pauvreté

A l'échelle locale

Concernant les milieux naturels

 Limiter l'appauvrissement des sols

 Protéger la faune

 Protéger la flore

 Préserver les paysages

 Limiter les bétonnages à outrance

Concernant les milieux humains

 Concernant la santé publique

 Limiter les effets de la pollution de l'air

 Des pics de pollutions

 De la pollution de fond

 Limiter les effets des accidents

 Limiter les effets du bruit sur la santé

 Faciliter l'accès aux soins

 Concernant la qualité de vie

 Limiter la gêne

 Due aux bruits

 Due aux fumées

 Due aux odeurs

 Favoriser les espaces verts

 Améliorer le paysage urbain

 Eviter les hausses de prix du foncier

 Offrir de l'art dans la rue

 Concernant la qualité des transports

 Faciliter la circulation individuelle

 Faciliter l'accès aux transports en commun

 Améliorer les services des transports en commun

 Faciliter les flux

Préserver le patrimoine culturel

Respecter les quartiers « villages »

Préserver les habitats des salissures

FIGURE 11 – Synthèse réduite

2.4 Retour d'expérience

Nous avons donc suivi la démarche proposés par Keeney [14] et nous nous sommes livrés à l'expérience de structurer des objectifs. Celle-ci a été riche. Tout d'abord, toutes les personnes rencontrées dans ce cadre ont manifesté un intérêt pour notre travail, ce qui a rendu notre tâche aussi agréable qu'intéressante. A posteriori, aucun commentaire n'a été apporté, ni de modification demandée par les personnes rencontrées suite à l'envoi des résultats. Il va de soi que les interlocuteurs ont accepté de se prendre au jeu sachant qu'ils n'impliquaient pas leur responsabilité à proprement dit compte tenu du caractère confidentiel de leur contribution et du caractère exploratoire de notre recherche.

Il n'y avait pas véritablement de pression au cours des rencontres. Les personnes interviewées ont facilement compris la consigne de départ. Notons que le fait d'appréhender le système de valeurs des interviewés en dehors de tout contexte réel (pas de projet concret, mais des considérations environnementales générales) a permis d'obtenir des discours semblant libre de toute contrainte, pression et autres facteurs susceptibles d'influencer les valeurs explicitées, ce qui va dans le sens de la théorie de Keeney ([14]).

La richesse des discours provient du fait que les personnes rencontrées avaient ou avaient eut des responsabilités dans le domaine des transports et/ou dans le domaine de l'environnement. Ces domaines leur étaient donc familiers. Dans l'avenir, il serait intéressant de reproduire une telle application (qui ne présente pas de réelle difficulté) car celle-ci date de 2003. Ce serait notamment l'occasion d'augmenter la taille de l'échantillon et d'obtenir ainsi un arbre actualisé et un peu plus exhaustif.

3 L'évaluation et l'agrégation

Après avoir structuré les objectifs sous forme d'arbre, il s'agit alors de rattacher à chacun d'entre un ou plusieurs critères pour les représenter. Or, lorsque l'ensemble des critères est important et de surcroît hétérogène, il convient assez naturellement d'opérer une agrégation multicritère. Cette partie est donc consacrée au problème de l'agrégation du résultat de l'évaluation. Nous rappelons tout d'abord le cadre de recherche dans lequel se positionne ce travail (section 3.1). Nous nous demandons ensuite comment utiliser l'agrégation multicritère pour combiner la problématique de tri avec celle de la description et finalement retenir une problématique spécifique à l'aide à l'évaluation (section 3.2). Nous proposons alors certaines orientations méthodologiques pour répondre au problème de l'agrégation multicritère (section 3.3) puis spatio-temporelle (section 3.4). Nous traitons ensuite de la pondération des critères (section 3.5). En suivant ces orientations, nous présentons dans la section 3.6.6 des exemples de modélisation (critère, catégories, seuils, agrégation spatio-temporelle) puis mettons en parallèle l'indice Atmo, bien connu dans le domaine de la santé et de la qualité de l'air, et la méthode multicritère ELECTRE TRI.

3.1 Contexte de l'évaluation

Besoin d'agrégation. Nous désirons contribuer à la conception de méthodes permettant d'aider les différents acteurs d'un processus de décision à se forger leur propre opinion sur l'état d'un système de nature complexe et le suivre au cours du temps. Or, cette tâche peut rapidement devenir complexe si ces acteurs ont des préoccupations nombreuses et variées concernant le système. Face à ce problème « multicritère », il est souvent difficile pour eux de se forger une opinion globale sur le système. Dès lors, l'aide à l'évaluation devient l'activité qui consiste à assister ses acteurs, notamment à l'aide de modèles, afin de leur permettre de mieux appréhender le système dans son ensemble. Pour tenter de palier cette complexité, les modèles utilisés doivent permettre (entre autre) de réduire les dimensions du problème en opérant certaines agrégations « multicritères ».

Echelles temporelles et spatiales d'observation. D'une part, le système que l'on souhaite évaluer est susceptible de changer au cours du temps et tandis que certains aspects peuvent évoluer très lentement, d'autres au contraire évoluent très vite et nécessitent par conséquent d'être observés plus souvent. D'autre part, le système peut présenter une dimension spatiale (il peut par exemple s'étendre géographiquement) et l'on se doit alors de considérer tous les sous-ensembles élémentaires qui composent l'ensemble géographique faisant l'objet de l'étude. Ces deux caractéristiques compliquent la tâche d'évaluation et, toujours dans un souci de réduire les dimensions du problème et ainsi aider à l'évaluation, il convient alors de proposer des modèles permettant des agrégations spatio-temporelles.

Approche par les objectifs. Comme nous l'avons vu en section 1.4.2, la question de l'évaluation d'un objet complexe repose avant tout sur l'observation de cet objet à la lumière d'un système de valeurs, c'est-à-dire d'un ensemble d'objectifs ou points de vue servant de référence. Nous considérons en effet que les décideurs possèdent a priori un certain nombre de préoccupations ou attentes concernant l'état du système à évaluer et qu'ils désirent alors confronter leur propre système de valeurs (constitué de ces préoccupations et attentes) à la réalité. Pour explorer le système de valeurs des décideurs il est possible de s'inspirer de la méthode de structuration des objectifs proposée par Keeney [14] et illustrée dans la partie 2. Cette méthode offre l'avantage d'être assez simple à appliquer et efficace pour recenser les attentes des décideurs en termes d'objectifs. Elle aboutit à la construction d'un « arbre d'objectifs » (section 2.1.7) synthétisant l'ensemble des points de vues exprimés. C'est cette approche que nous retenons comme une étape préalable.

3.2 Problématique de l'aide à l'évaluation

Nous avons vu en section 1.5.3 que deux problématiques d'aide à la décision semblent avoir un intérêt en aide à l'évaluation : celle de tri et celle de la description. Tandis que la première opère une agrégation multicritère, ce n'est pas le cas de la seconde. Or, l'agrégation multicritère contribue tant à introduire de la clarté dans la présentation des résultats, rendus de fait plus condensés et synthétiques, qu'à introduire de l'opacité à cause des règles, méthodes, paramètres, que l'agrégation utilise pour synthétiser les résultats. Agréger contribue donc à deux phénomènes antagonistes : elle contribue à la lisibilité des résultats et, en même temps, elle nuit à la transparence de ces mêmes résultats.

Il semble nécessaire, dans notre contexte exigeant à la fois une certaine transparence et la simplicité du résultat (Cf. Introduction), d'éviter de tomber dans l'un de ces deux écueils. Pour cela, nous proposons de présenter à la fois des résultats agrégés et désagrégés. C'est une façon de proposer une problématique combinant celle de tri et celle de la description.

Notons à cet effet que les travaux de Moussa [19] proposent une application utilisant une approche d'agrégation-désagrégation fondée sur un modèle hiérarchique multicritère. Dans cette approche, l'agrégation est réalisée à différents niveaux de l'arbre et peut donc concerner un sous-ensemble des critères identifiés. A l'exception de ce travail, les critères à agréger en aide à la décision pour une problématique de tri (qui nous intéresse tout particulièrement) sont ceux de l'ensemble de la famille F de critère [25] et il ne semble pas pertinent de réaliser une agrégation de critères d'un sous-ensemble de cette famille. En effet, l'agrégation se doit de prendre en compte tous les critères pour offrir un résultat pourvu d'une signification dans une optique d'aider à prendre une décision. Ecarter certains critères reviendrait à considérer seulement certaines des conséquences d'une décision et contribuerait à appauvrir la signification du résultat agrégé qui n'aurait alors plus un sens global vis-à-vis du système de valeurs du décideur. Autrement dit, nous pensons qu'il n'est pas souvent pertinent d'aider à décider en ne considérant (en n'agrégeant) qu'une sous partie des critères identifiés.

L'aide à l'évaluation proposée en section 1 s'appuie aussi sur un arbre d'objectifs ce qui n'est pas systématiquement le cas de l'aide à la décision. Rappelons que, dans notre approche, chaque critère est rattaché à un objectif de plus bas niveau, un objectif non décomposable (une feuille de l'arbre). Prenons alors un objectif de l'arbre qui ne soit pas une feuille. Cet objectif engendre un sous-arbre dont il est la racine. De fait, les critères rattachés aux différentes feuilles de ce sous-arbre forment un ensemble relativement cohérent par rapport à la signification de l'objectif considéré. Agréger ces critères semble avoir une signification : cela permet d'obtenir une évaluation du système (sous forme du résultat d'un tri dans notre contexte) du point de vue de l'objectif racine en question.

On peut alors imaginer autant d'agrégations possibles (et finalement autant de résultats d'évaluation) que d'objectifs dans l'arbre. Autrement dit, nous proposons, à chaque nœud, d'affecter le système évalué à une catégorie. Cette affectation est réalisée en prenant en compte tous les critères issus de la décomposition de cet objectif.

Cette façon d'évaluer à chaque nœud de l'arbre semble offrir un certain continuum entre la problématique de tri et de la description. En effet, cela offre à l'évaluant la possibilité de parcourir l'arbre d'objectifs en consultant à chaque nœud le résultat agrégeant tous les critères directement ou indirectement rattachés à l'objectif en question. En partant par exemple de la racine de l'arbre, l'évaluant pourra alors naviguer en descendant (éventuellement jusqu'aux objectifs non décomposables) et remontant à sa guise tout en consultant à chaque étape un résultat agrégé, une évaluation.

Une telle fonction semble pallier l'antagonisme entre simplicité et transparence du résultat en rendant possible, lors de la consultation des résultats, l'alternance de phases d'agrégation de l'information (remontée dans l'arbre des objectifs) et de désagrégation (descente). C'est cette problématique que nous retenons pour l'aide à l'évaluation dont il est question dans ce travail.

3.3 Agrégation multicritère

Les méthodes d'agrégation multicritère existantes ont été développées principalement dans le but d'aider un acteur à prendre une décision. En section 1.5.3, nous avons retenue une problématique spécifique inspirée de la problématique de tri et de la problématique de la description (problématiques issues de l'aide à la décision). Celle de Tri nécessite une agrégation multicritère. Or, pour agréger, il existe deux grandes familles de méthodes que nous allons maintenant présenter puis comparer au regard de notre contexte.

3.3.1 Méthodes d'agrégation utilisant un critère unique de synthèse

Les études relatives à l'évaluation comparative d'action utilisent souvent une agrégation qui ramène un ensemble d'indicateurs par nature hétérogènes et disparates à une dimension unique, ou critère de synthèse. Il s'agit dans la majorité des cas d'une dimension monétaire (ou d'utilité). Les différentes conséquences sont toutes traduites en termes de coût ou de bénéfices. L'évaluation globale consiste alors à faire le bilan monétaire de l'ensemble des coûts et des bénéfices induits. Comparer les différentes actions se réduit à comparer les valeurs de leur bilan monétaire. Les approches moné-

taires restent les plus utilisées, bien qu'elles soient souvent critiquées car s'il est clair que certaines conséquences sont aisément convertibles en termes monétaires sans que cela n'induisse trop d'arbitraire, c'est bien moins évident pour d'autres. Au-delà même de cette conversion en termes monétaires, de nouveaux problèmes apparaissent lors de l'étape de synthèse qui consiste à faire la somme des coûts et des avantages ([4]). En effet, faire le bilan sur une même échelle monétaire de conséquences aussi variées que des aspects environnementaux, sociaux, humains ou organisationnels pose problème, notamment du fait de compensations des coûts et avantages. Est-il acceptable de laisser pleinement se compenser des conséquences dans des domaines aussi variés ? Or, les méthodes conduisant à établir un critère unique peuvent difficilement s'affranchir de cette pleine compensation. En théorie, la mise en plac de ces méthodes nécessite un lourd travail d'élicitation des préférences. Dans la pratique, on occulte souvent ces étapes de modélisation et l'on manipule le résultat final sans précautions. En conclusion, ramener un ensemble de conséquences hétérogènes sur une échelle commune présente l'avantage d'un résultat numérique, facile à appréhender et à exploiter, mais d'une part la compensation ne peut être maîtrisée dans ces méthodes monétaires, d'autre part leur mise en œuvre rigoureuse est fastidieuse et de ce fait négligée. En conséquence, si ces méthodes paraissent assez adaptées pour évaluer certains aspects (coûts, bénéfiques), elles semblent l'être beaucoup moins pour évaluer les facettes environnementales, sociales, humaines ou organisationnelles.

3.3.2 Méthodes d'agrégation de type ELECTRE

Parmi les méthodes permettant l'agrégation multicritère, nous décrirons plus précisément celles de type ELECTRE [25], lesquelles sont fondées sur des comparaisons par paires d'actions et sur la notion de « surclassement ». Dans ces méthodes, on compare par exemple deux actions de façon globale, pour aboutir à une des trois conclusions suivantes :

- l'une est meilleure que l'autre (sans préciser l'écart),
- les actions sont incomparables,
- les actions sont indifférentes.

Fondées sur l'utilisation du formalisme algébrique de la théorie des ensembles, les comparaisons sont réalisées sans essayer de ramener les différentes dimensions sur une même échelle de valeurs. Elles utilisent des indices de « concordance » et de « discordance ». Ces deux d'indices ont été construits en vue d'établir ou non le surclassement d'une action par une autre en testant les deux conditions suivantes :

- il y a bien majorité suffisante de critères en faveur du surclassement (indice de concordance),
- parmi la minorité, aucun critère ne s'oppose trop fortement au surclassement (indice de discordance).

Ces indices sont notamment calculés à partir de poids (reflétant un « pouvoir de vote ») que l'on a choisi de conférer à chaque critère et de seuil de veto (reflétant un « pouvoirs de veto ») conférés à certains critères.

Les concepts sous-jacents à ce type de méthodes (seuils, principe utilisé pour l'agrégation et autres) paraissent bien adaptés à la plupart des domaines pour lesquels les critères sont hétérogènes et/ou entachés d'imprécision et d'incertitude et nous préconisons en conséquence de les utiliser dans ces cas. La possibilité offerte par ELECTRE de fixer des seuils de veto est un moyen supplémentaire de maîtriser le danger de la compensation en sanctionnant des écarts trop importants, même s'ils concernent des critères de faible poids. Le système relationnel de préférence sous-jacent à la construction de critères permet de manière simple, notamment grâce à l'introduction de seuils d'indifférence et de seuils de préférence, de traiter aussi les données de mauvaise qualité en distinguant dans la modélisation les écarts pertinents et significatifs de ceux qui ne le sont pas.

Malgré tous ces avantages, la pratique de telles méthodes est encore marginale. Les résultats obtenus sont certes moins faciles à exploiter qu'un critère unique de synthèse, puisque le modèle ne fournit pas d'écart entre deux actions et qu'il aboutit même parfois à considérer deux projets comme incomparables, là où les méthodes monétaires auraient tranché. Ce dernier point peut cependant être considéré plutôt comme un avantage de ces méthodes, car nous sommes confrontés parfois réellement à de telles situations. Ces méthodes d'agrégation sont très intéressantes pour comparer différentes perspectives ou simulations, par exemple plusieurs projets concurrents. Plus précisément, elles ont été développées pour répondre à trois des quatre problématiques présentées en section 1.5.1 (sélection, rangement et tri). Cependant, il est souvent nécessaire de reconstruire tout ou partie du modèle à chaque nouvelle problématique décisionnelle, si celle-ci se singularise des précédentes par sa nature, par ses objectifs ou par ses acteurs. Or, dans notre contexte, on cherche avant tout à évaluer un système complexe par comparaison à une *norme* (sorte de système de référence). La problématique de tri, qui consiste à affecter l'action (dans notre cas, le système) à une catégorie prédéfinie, est bien adaptée à ce cas. À cet effet, nous proposons d'utiliser la méthode ELECTRE TRI [25].

3.4 Agrégation spatio-temporelle

Notre contexte décrit en section 3.1 nécessite une agrégation spatiale et temporelle. Dans un autre contexte, cela ne semblerait pas forcément approprié. Cependant, on peut imaginer des applications qui nécessiteraient l'agrégation de dimensions autres que le temps ou l'espace comme agréger les opinions des individus d'une population. Nous allons discuter de la façon de prendre en compte une dimension supplémentaire pour un critère avec l'exemple du temps et de l'espace.

Tandis que certains aspects peuvent évoluer très lentement (réchauffement climatique), d'autres au contraire évoluent très vite et nécessitent par conséquent d'être observés plus souvent (pics de pollution). Pour choisir une échelle temporelle d'observation commune à deux aspects n'ayant pas la même fréquence d'observation, il s'agit par exemple de construire un critère annuel à partir des 365 valeurs journalières. Pour cela, une première technique serait de réaliser une agrégation temporelle en ramenant ces 365 valeurs en un critère unique de synthèse grâce par exemple à des méthodes de moyennes. De plus, certains critères sont parfois observables dans plusieurs sous-ensembles de l'ensemble géographique étudié (relevés des pics de pollution à divers endroits d'une ville).

Se pose alors le problème de l'agrégation spatiale des différentes valeurs observées. Dans ce cas aussi, il semble a priori aisé de réaliser un moyennage ou une autre opération de même type.

Cependant, au lieu d'agréger systématiquement plusieurs mesures (de même nature) par une méthode se ramenant à un critère unique de synthèse, nous pensons qu'il est aussi possible et intéressant de conserver chacune des observations et de les utiliser dans une procédure d'agrégation de type ELECTRE. Chacune d'elle pourra alors participer au vote (dans le calcul de l'indice de concordance pour établir le surclassement) et aura, le cas échéant, possibilité d'appliquer son veto (dans le calcul de l'indice de discordance pour s'opposer au surclassement).

Cette manière de tenir compte de la présence de dimensions supplémentaires consiste finalement à répliquer le critère en question autant de fois que la dimension le nécessite (1 critère annuel issu de 365 critères journaliers). Selon le formalisme d'ELECTRE, cela revient à transformer un critère en plusieurs critères (autant que de mesures disponibles dans la dimension) de mêmes caractéristiques d'un point de vue de leur modélisation (échelle, seuils, catégories, etc) mais qui n'auront a priori pas la même performance (la même évaluation). Dans les deux cas passés en revue (mesures ponctuelles dans l'espace ou dans le temps), se pose alors la question du pouvoir de vote, du poids de chacune des mesures. Une condition qui nous semble raisonnable est que la somme des poids de toutes les mesures doit coïncider avec le poids accordé au critère. Le poids accordé à l'objectif initial est ainsi réparti entre tous les critères élémentaires. La façon de répartir ce poids entre les différentes valeurs élémentaires peut être variée. On peut imaginer répartir ce poids de manière équitable entre toutes les valeurs élémentaires. On peut aussi définir une règle de répartition qui tiendrait compte d'autres paramètres reflétant l'importance d'une observation relativement aux autres.

Nous pensons en tout cas que cette façon de procéder est intéressante pour au moins trois raisons :

- éviter de tomber dans les écueils classiques des moyennes et autres modes de calculs semblables,
- tenir compte d'autres paramètres pour répartir le poids du critère entre les différentes observations spatio-temporelles,
- conférer un pouvoir de veto à toute mesure trop mauvaise, aussi noyée soit elle dans un ensemble de mesures meilleures.

Cette approche nous paraît être une alternative intéressante pour éviter une agrégation préalable de la dimension avec des méthodes de type critère unique de synthèse (moyenne, somme pondérée, méthode statistique de lissage etc). En effet, cela permet de conserver le détail de l'information contenu dans les différentes valeurs ponctuelles par le biais du pouvoir de vote dont dispose alors chacune d'elle. Théoriquement, cela semble donner l'avantage d'un vote moins tranché et donc plus nuancé dans le sens où il sera rare que tout le poids accordé à l'objectif vote pour ou contre le surclassement. Le choix du jeu de poids ainsi que la valeur du paramètre de la majorité semblent alors moins sensible dès lors que le problème d'évaluation nécessite l'ajout d'une dimension. Enfin, le pouvoir de veto éventuellement conféré à certaines de ces valeurs leur permet

de s'opposer à elles seules à une majorité. En conséquence, contrairement aux méthodes de type critère unique de synthèse, cela permet d'éviter de noyer certaines valeurs exceptionnellement mauvaises dans un ensemble de bonnes valeurs. Cependant, une limite de cette méthode peut apparaître lorsqu'il existe une certaine dépendance entre les valeurs observées ponctuellement. Dans l'exemple de la dimension temporelle, il est reconnu que les séries chronologiques montrent parfois des composantes saisonnières. De tels phénomènes compliquent les choses.

Finalement, la façon de répondre au problème de l'agrégation géographique et temporelle proposée semble simple et avantageuse. Elle mériterait à ce titre d'être considérée comme une voie intéressante pour traiter le problème de l'agrégation spatio-temporelle. Notons à cet effet que Chakhar [7] propose une manière plus fine de remédier à ce problème en vue d'une application à l'aménagement des infrastructures linéaires et en particulier dans un contexte décisionnel de choix de corridors.

3.5 Pondération des critères

La problématique d'aide à l'évaluation que nous avons retenue en section 3.2 propose, à chaque nœud de l'arbre d'objectifs, d'opérer une agrégation de tous les critères issus de la décomposition de l'objectif en question. Or, l'agrégation multicritère soulève très généralement (c'est le cas dans ELECTRE TRI) la question de la pondération des critères. Nous avons vu que la question de la pondération des objectifs est du ressort de l'évaluant. Il s'agit en effet d'un jugement de valeurs qui consiste à attribuer de l'importance aux différents critères par l'intermédiaire d'un pouvoir de vote (au sens d'ELECTRE). Dans la pratique de l'aide à la décision, il existe une méthode dite « méthode des cartes » révisée par Roy et Figueira [26] qui permet de déterminer un jeu de poids (utilisable avec les méthodes ELECTRE) pour une famille de critères donnée. Une analyse critique de cette méthode est proposée dans [5] où des alternatives sont proposées. Voici cependant son principe.

Méthode des cartes. Chaque critère est représenté par une carte marquée de son nom. C'est une méthode manuelle qui permet de bien se représenter l'importance relative des critères. Elle aboutit à attribuer à chaque critère de la famille F , un poids représentant l'importance que le décideur lui accorde relativement aux autres critères de F . La première étape consiste à regrouper les critères équivalents au sens des préférences des décideurs en paquets, s'il y a lieu. À partir de ces paquets et des cartes restantes, le décideur doit en second lieu, classer les cartes par ordre de préférence. Ensuite, il doit choisir d'introduire des cartes vierges (symbolisant l'écart de préférence) entre chacun des paquets ou cartes seules. On obtient donc une pile de cartes et paquets de cartes. Une fois cela réalisé, le décideur doit évaluer numériquement de combien le critère le plus important (sommet de la pile) est plus important que le critère le moins important (bas de la pile). À partir de ce travail de modélisation, la méthode permet de trouver un jeu de poids correspondant aux préférences des décideurs.

Si, en aide à la décision les décideurs doivent pondérer tous les critères de la famille F relativement les uns aux autres, ici, il s'agit de pondérer les objectifs rattachés à un

même nœud de l'arbre. Le travail de pondération concerne donc un sous-ensemble de la famille F ce qui le rend d'autant moins compliquer que ce sous-ensemble est restreint. En revanche, on est obligé de réitérer cette étape de pondération autant de fois qu'il y a de nœuds) dans l'arbre d'objectifs ce qui peut sembler d'autant plus fastidieux que cet arbre est grand.

Finalement, nous proposons une démarche qui consiste à fractionner l'étape de pondération des critères. Elle s'appuie sur l'arbre d'objectifs et se décompose en deux étapes. Pour la détailler, nous noterons $S(O)$ l'ensemble des suivants de l'objectif O , c'est-à-dire l'ensemble des sous-objectifs directement rattachés à l'objectif O . Notons ensuite P_0 l'objectif précédent de O , c'est-à-dire l'objectif plus général auquel est directement rattaché l'objectif O . Si $S(O)$ contient en général plusieurs objectifs (tous les suivants de O), P est unique car dans un arbre un objectif n'a qu'un seul précédent direct. Les étapes sont alors les suivantes :

- pour chaque nœud N de l'arbre (hormis les feuilles), définir un jeu de poids pour l'ensemble $S(N)$,
- pour une agrégation souhaitée à un nœud donné, tenir compte des jeux de poids pour déterminer le poids chaque objectifs de plus bas niveau (les feuilles) auxquels sont rattachés les critères à agréger.

La première étape est nommée « pondération des nœuds de l'arbre » et la seconde « propagation des poids ».

3.5.1 Pondération des nœuds de l'arbre

Pour chaque nœud N de l'arbre l'évaluant doit alors pondérer chacun des éléments de $S(N)$. Il doit finalement attribuer à chaque objectif $O_i \in S(N)$ un poids noté n_{O_i} . Pour ce faire, on peut préconiser la « méthode des cartes » décrite précédemment. Notons que si l'arbre est étendu, ce travail de pondération peut devenir conséquent, car il doit être fait à chaque nœud de l'arbre des objectifs.

Nous souhaitons, d'un point de vue pratique et pour préparer l'étape suivante de « propagation des poids » dans l'ensemble de la hiérarchie, normaliser chaque jeu de poids. Pour chaque $O_i \in S(N)$ nous retiendrons finalement un poids normalisé noté e_{O_i} et défini de la façon suivante :

$$e_{O_i} = \frac{n_{O_i}}{\sum_{O_j \in S(N)} n_{O_j}}$$

Ainsi les e_{O_i} sont des nombres compris au sens large entre 0 et 1 et à chaque nœud N on a :

$$\sum_{O_j \in S(N)} e_{O_j} = 1$$

3.5.2 Propagation des poids

Il reste maintenant à définir la façon dont on va se servir de ces différentes pondérations pour préparer l'agrégation à un nœud M donné. A priori, il n'y a pas une seule

façon de procéder mais nous en proposons une pleinement compatible avec l'utilisation d'ELECTRE TRI.

Considérons que $N \in S(M)$. Autrement dit, N est un sous-objectif directement rattaché à M . Soit e_N le poids normalisé attribué à l'objectif N . e_N représente l'importance relative qu'apporte l'évaluant à l'objectif N par rapport aux autres objectifs de $S(M)$.

Intéressons nous maintenant à un objectif $O \in S(N)$. En fait, O est rattaché à N qui lui même est rattaché à M . Notons $e_O(M)$ le poids de l'objectif O dans une agrégation multicritère au niveau M . Nous souhaitons définir $e_O(M)$ en tenant compte du poids e_O que possède O relativement aux autres objectifs de $S(N)$, mais aussi en tenant compte du poids e_N que possède N par rapport aux autres objectifs de $S(M)$. Nous proposons la définition $e_O(M) = e_O \times e_N$. De cette manière, le poids e_N accordé à N au niveau supérieur va se répartir entre tous les objectifs de $S(N)$ proportionnellement au poids de chacun d'entre eux. Nous remarquons que, grâce à la normalisation :

$$\sum_{O_j \in S(N)} (e_{O_j} \times e_N) = e_N \times \sum_{O_j \in S(N)} e_{O_j} = e_N$$

Toujours dans l'optique d'une agrégation multicritère souhaitée au niveau M , nous allons maintenant généraliser le principe de propagation afin de déterminer le poids d'un objectif O quelconque de l'arbre. O doit faire partie du sous-arbre dont M est la racine sans quoi O ne serait pas pris en compte dans l'agrégation. On alors peut définir le poids de O de la manière suivante :

$$e_O(M) = e_O \text{ si } O \in S(M)$$

$$e_O(M) = e_{P_O}(M) \times e_O \text{ sinon.}$$

où $e_{P_O}(M)$ est le poids de P_O dans l'agrégation souhaitée au niveau M .

Il y a un caractère récursif dans cette définition car elle fait appel au poids $e_{P_O}(M)$ de l'objectif P_O (objectif auquel O est directement rattaché). En conséquence, pour déterminer tous les poids nécessaires à l'agrégation souhaitée au nœud M , il faut opérer de manière descendante, c'est-à-dire commencer par déterminer les poids des suivants de M puis ceux des suivants des suivants de M et ainsi de suite jusqu'aux feuilles.

3.5.3 Niveau de majorité

Dans ELECTRE TRI, fixer la majorité requise (λ) c'est modéliser une information intercritère. Plus λ est proche de 100%, plus le modèle est sévère pour établir un surclassement et plus il risque d'apparaître la situation d'incomparabilité lors de la comparaison de deux actions. Dans notre contexte d'aide à l'évaluation, nous avons proposé de rendre possible l'agrégation multicritère de type ELECTRE à tous les niveaux de l'arbre d'objectifs, ou plus précisément au niveau de chaque objectif décomposable en sous-objectifs, à chaque nœud de cet arbre. À cet effet, il n'y a pas de raisons particulières qui nous pousseraient à fixer le paramètre λ de façon unique pour tous les nœuds de l'arbre. On

envisage donc la possibilité pour l'évaluant de nuancer la sévérité qu'il impose au modèle selon l'endroit où il se positionne dans l'arbre. Cela revient donc à envisager un paramètre λ pour chaque nœud.

Le paramètre λ doit être fixé au regard du jeu de poids retenu pour l'ensemble des critères de la famille F . En effet, prenons l'exemple où seulement trois critères forment la famille F , l'un de poids 0,7 et les deux autres de valeur 0,15. Si on fixe $\lambda = 0,7$ cela revient à donner plein pouvoir au critère le plus pondéré pour établir ou non le surclassement. En revanche, si on fixe λ à 0,8 alors ce critère à lui seul n'est plus apte à établir le surclassement. Pour fixer ces valeurs de λ , nous proposons à l'évaluant de s'appuyer sur les valeurs des poids issues de l'application de la méthode des cartes décrite ci-avant. Il faut pour cela combiner les différentes valeurs de poids afin d'étudier les coalitions de critères qui suffisent (ou ne suffisent pas) à constituer une majorité, cela en vue de vérifier que la modélisation retenue (poids et paramètre λ) est bien fidèle au système de valeurs de l'évaluant.

Dans ELECTRE, un critère vote la plupart du temps (hormis dans le cas de la préférence faible) de manière tranchée, c'est à dire soit avec tout son poids, soit au contraire avec aucun poids. C'est l'une des raisons pour laquelle le paramètre λ doit être fixé au regard du jeu de poids de la famille F . Rappelons que le principe de propagation des poids décrit en section 3.5.2 conduit en quelque sorte à diffuser les poids (ou pouvoirs de vote) dans l'arbre d'objectifs. Le vote d'un objectif dépend alors des votes de ses sous-objectifs qui n'iront a priori pas tous dans le même sens. Ajoutons à cela que nous avons proposé une manière de traiter le problème de l'agrégation spatio-temporelle qui consiste à répliquer un critère autant de fois qu'il y a d'observations disponibles dans l'espace et/ou dans le temps. Chaque « réplique » du critère se voit alors investi d'une partie du poids initial. Ces deux orientations méthodologiques (principe de propagation et solution proposée pour gérer les dimensions spatio-temporelles) conduisent de fait à des votes beaucoup moins tranchés. Ainsi, dans notre cas, le paramètre λ est moins sensiblement lié au jeu de poids de la famille F .

3.6 Exemple de modélisation

Nous allons présenter une partie de la modélisation d'ELECTRE TRI proposée par Rousval [22] dans le contexte d'une évaluation environnementale des transports sur l'environnement. Dans ce contexte particulier, le *système évalué* est un système de transports et la situation environnementale qui en découle.

3.6.1 Le critère

Dans la section 2.3.3, en appliquant l'approche de structuration des objectifs proposée par Keeney [14], nous avons obtenu un objectif de santé publique intitulé « Maîtriser l'environnement - A l'échelle locale - Concernant les milieux humains - Concernant la santé publique - Limiter les effets de la pollution de l'air - Des pics de pollutions » (voir figure 11). Il concerne les polluants dans l'air et leurs impacts sur la santé. Pour notre exemple de modélisation, nous allons définir un critère concernant un polluant qui a déjà

donné lieu à un travail de recherche conséquent : le monoxyde de carbone CO. C'est en effet l'un des rares exemples de gaz polluant et toxique pour lequel on a pu établir une relation (différentielle) entre la teneur de gaz respiré et la composition du sang du sujet exposé (Chovin et Richalet [8]) et l'on en a déduit un abaque (figure 12) illustrant la correspondance entre une exposition de quelques heures et ses conséquences sur la santé classées en six catégories ([27], [15, 17]). Il s'agit donc d'une description précise d'impact qui se révèle intéressante et nous avons repris ce cas, même si les progrès technologiques ont fortement atténué le problème du CO émis par les véhicules particuliers. Un critère pourrait donc être la teneur en CO dans l'air (exprimée en ppm, « partie par million ») qui contribuerait à mettre en exergue le danger pour la santé de pics de pollution d'un certain type.

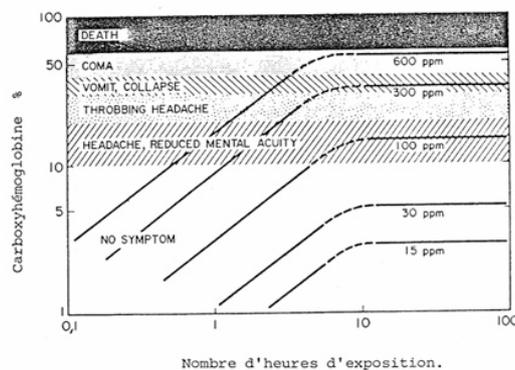


FIGURE 12 – Lien entre l'exposition au CO et les effets sur la santé 12

3.6.2 Catégories

Pour appliquer ELECTRE TRI, il faut découper les échelles de tous les critères en catégories ordonnées. Les limites séparant les catégories sont appelées *profils limites* [25]. Précisons que la définition des profils limites est normalement de la responsabilité de l'évaluateur car il s'agit de *normes* dont dépend le résultat de l'évaluation : cela peut être vu comme un certain niveau d'exigence. Cependant, lorsque le critère est très technique, l'évaluateur peut bien entendu être assisté par les experts dans cette tâche. Pour cela, il faut préalablement fixer le nombre de catégories que l'on désire constituer. À cet effet, nous nous sommes inspirés du protocole des échelles Atmo [2], qui introduit trois découpages en catégories :

- un premier découpage en 10 catégories définissant la valeur de l'indice allant de 1 à 10 (les situations les plus favorables pour l'environnement étant classées dans la catégorie d'indice le plus faible et les plus mauvaises dans la catégorie d'indice le plus élevé),
- un second découpage en 6 catégories qualifiant l'état du système (très bon, bon, moyen, médiocre, mauvais et très mauvais),

- un dernier découpage en 3 couleurs (vert, orange et rouge).

Ces trois découpages sont emboîtés les uns dans les autres. Les catégories sont définies à l'aide de valeurs limites (équivalentes aux profils limites d'ELECTRE), respectivement b_i, c_i et d_i , valeurs qui ont fait l'objet de discussions entre spécialistes des nuisances et résultent d'un consensus.

Pour les échelles des critères que l'on découpe en 10 catégories, les valeurs limites vont de b_1 à b_9 . Pour les échelles des critères que l'on découpe en 6 catégories, les valeurs limites vont de c_1 à c_5 . Pour les échelles des critères que l'on découpe en 3 catégories, les valeurs limites sont d_1 et d_2 . Pour permettre les emboîtements, il y a correspondance entre les valeurs limites respectives : en l'occurrence $b_2 = c_1$; $b_4 = c_2 = d_1$; $b_5 = c_3$; $b_7 = c_4 = d_2$ et $b_9 = c_5$. Précisons que, selon la nature des critères, il est possible de définir soit les b_i soit les c_i soit les d_i .

A partir de ces options de modélisation, tentons de définir les catégories pour notre exemple de critère. Les effets sur la santé dus à l'exposition au CO sont relativement bien connus [15]. Ces effets dépendent du taux dans l'air respiré ainsi que de la durée d'exposition à ce taux. Notons cependant qu'après environ 10 heures d'exposition à un même taux, les effets n'évoluent plus.

En exploitant la figure 12 nous pouvons donc établir les tranches suivantes pour de longues expositions (près de 10 heures) :

- pas de danger pour les expositions à des taux inférieurs à 60 ppm,
- maux de tête et baisse de l'acuité mentale pour des taux compris entre 60 et 120 ppm,
- forte amplification du phénomène entre 120 et 250 ppm,
- vomissement et évanouissent entre 250 et 350 ppm,
- coma pour des taux au-delà de 350 ppm,
- mort au-delà de 600 ppm.

Ces valeurs sont un peu approximatives car elles sont issues de lectures graphiques, cependant, elle reflètent bien l'évolution du phénomène en fonction de la teneur en CO dans l'air. Les approximations graphiques ont été réalisées de manière prudente, c'est-à-dire que les valeurs ont été arrondies par défaut. Pour donner une signification à ces catégories, nous proposons de qualifier de « très mauvais » les trois derniers états considérés. Autrement dit, à partir de l'état de vomissement et de perte de connaissance, la situation est très mauvaise. On peut ainsi suggérer les valeurs limites des catégories suivantes :

- très bon, teneur entre 0 et 10 ppm,
- bon, teneur entre 10 et 25 ppm,
- moyen, teneur entre 25 et 45 ppm,
- médiocre, teneur entre 45 et 70 ppm,
- mauvais, teneur entre 70 et 120 ppm,
- très mauvais, teneur au-delà de 120 ppm.

Techniquement, pour construire ces six catégories sur le critère « teneur en CO en ppm », cinq profils limites c_i sont nécessaires :

- $c_1 = 10$ ppm,

- $c_2 = 25$ ppm,
- $c_3 = 45$ ppm,
- $c_4 = 70$ ppm,
- $c_5 = 120$ ppm,

3.6.3 Seuils de discrimination

Les six catégories définies pour ce critère ne sont pas de même étendue (10 ppm pour la première, 50 ppm pour la cinquième). D'ailleurs l'étendue des catégories augmente avec les teneurs. Pour cette raison mais aussi pour le caractère progressif de l'évolution des conséquences sur la santé nous optons ici pour des seuils (préférence et indifférence) variables. Nous pensons qu'il est raisonnable de fixer le seuil d'indifférence à 5% de la teneur et le seuil de préférence à 10% de la teneur. Autrement dit, on est indifférent entre 10 et 10,5 ppm en revanche on préfère 10 ppm à 11 ppm. De même, on est indifférent entre une teneur de 120 et une teneur jusqu'à 126 ppm, mais on préfère nettement une teneur de 120 à une teneur de 132 ppm. Ces seuils permettent aussi de rendre compte de l'approximation faite lors de la lecture graphique pour fixer les valeurs des profils limites. Nous précisons que le travail de définition des catégories et des seuils est bien entendu du ressort d'experts (ici, experts de la santé). Nous ne prétendons donc pas avoir fixé ces valeurs de façon la plus précise ni la plus juste. Cette modélisation est proposée avant tout à titre d'illustration.

3.6.4 Seuil de veto

Selon Roy et Bouyssou [25], le principe d'un seuil de veto est de conférer à un critère, le pouvoir de s'opposer à lui seul au surclassement quelque soit le poids qu'on lui a accordé. Dans le cadre d'ELECTRE TRI, cela revient à donner au critère le pouvoir de s'opposer à l'affectation d'un système à une catégorie. Dans notre application, il s'agit de modéliser une fonction d'alerte par le biais de ce seuil. Par exemple, même si le système est « très bon » sur tous les critères sauf sur la teneur en CO où il est « mauvais », alors on peut conférer un pouvoir de veto au critère CO pour qu'il s'oppose à l'affectation globale du système dans la catégorie « très bon », « bon », voire « moyen ». En fixant le seuil de veto à 60 ($b_7 - b_2$) sur ce critère, une teneur de 110 ppm empêcherait d'affecter globalement le système à la catégorie « moyenne », « bonne » ou « très bonne » car quelles que soient les performances du système sur les autres critères pour les profils limites inférieurs de ces trois catégories on vérifie alors la condition de veto. En effet, $110 - 10 > 60$, $110 - 20 > 60$, et $110 - 45 > 60$.

En revanche, avec cette modélisation, le critère CO ne s'opposera pas à l'affectation du système aux autres catégories (comme par exemple la catégorie « médiocre ») même si du point de vue CO le système est « mauvais » avec une teneur de 110. Nous proposons de conserver la valeur 60 pour seuil de veto sur le critère « teneur en CO » pour plusieurs raisons. D'une part, parce que les effets du CO sur la santé peuvent être catastrophiques, contrairement aux nuisances de gênes. D'autre part, car la valeur 60 permet pour une teneur de 120 ppm (système très mauvais) d'empêcher l'affectation globale du système

à la catégorie « vert », et même « moyen ». Pour des teneurs encore plus élevées et à partir de 130 ppm, le système sera affecté à « rouge ». Pour des teneurs en dessous de 70 ppm, ce veto ne jouera plus aucun rôle. Finalement, ce seuil permettra de couvrir une fonction d’alerte concernant la dangerosité due à la présence trop importante de CO dans l’air (situation rouge).

3.6.5 Principe d’agrégation spatio-temporelle

Tout d’abord, considérons le cas où plusieurs observations sont réalisées en des différents lieux du territoire considéré. Se pose alors la question du pouvoir de vote (poids) de chacune des observations (les différentes teneurs en CO observées). Une condition qui nous semble raisonnable est que la somme des poids de toutes les observations doit coïncider avec le poids accordé au critère. Le poids initial est ainsi réparti entre tous les critères élémentaires, entre toutes les observations réalisées sur le territoire considéré. On peut imaginer de répartir simplement ce poids de manière équitable entre toutes les valeurs élémentaires. On peut aussi définir une règle de répartition du poids qui tiendrait compte d’un autre paramètre. Si l’on considère que les appareils de mesure des teneurs en CO sont placés dans des lieux de même importance, en suivant la règle d’agrégation spatio-temporelle proposée en section 3.4, on peut alors répartir le poids du critère équitablement entre les différentes observations. Cependant, si l’on considère que certains appareils de mesure sont situés dans des lieux plus fréquentés que d’autres (là où la population exposée à la pollution est plus grande) on peut alors apporter plus d’importance aux valeurs relevés par ces appareils que par les autres. Il s’agit pour l’expert en pollution de trouver une clé de répartition qui symbolise l’importance relative de chacune des teneurs observées. Cette clé peut être fixe, mais on peut aussi imaginer qu’elle soit issue de relevés de fréquentation des lieux où sont situés les appareils de mesure.

Ensuite, remarquons que notre critère peut aussi nécessiter une agrégation temporelle si des observations successives sont réalisées durant la période considérée (la journée par exemple). Cela ne pose a priori pas de problème de combiner à la fois une première répartition du poids du critère entre les différents appareils de mesure, puis le poids de chaque appareil de mesure entre les différentes valeurs relevées par cet appareil durant la période considérée (la journée).

3.6.6 Le cas de l’indice Atmo

Prenons l’exemple de l’indice Atmo dont une bonne description est faite par Garcia et al. [11]. Il s’agit d’un indice calculé obligatoirement chaque jour dans toutes les agglomérations françaises de plus de 100 000 habitants par les associations agréées de surveillance de la qualité de l’air (AASQA). Cet indice est donc opérationnel. Une station de mesure fournit pour un polluant une valeur journalière qui peut être soit la moyenne de toutes les valeurs de la journée (cas des particules) soit la moyenne des maxima horaires (cas du dioxyde de soufre SO_2 du dioxyde d’azote NO_2 et de l’ozone O_3). Sur chaque polluant, l’indice journalier pour la ville est alors calculé à partir des relevés réalisés par au minimum deux capteurs situés en zone urbaine ou périurbaine, et choisis de telle façon que la

moyenne des mesures réalisées par ces stations soit représentative des concentrations urbaines. La valeur de l'indice Atmo est alors le maximum entre les quatre valeurs d'indices prises sur chacun des polluants. Pour une diffusion le jour même, les moyennes de toutes les valeurs journalières ne pouvant pas être effectuées, il s'agit alors d'une estimation de l'indice à 16h, estimation pouvant être complétée par une prévision de l'indice pour la journée entière. Le choix des différentes valeurs limites ont permis d'aboutir à un indice sensé refléter l'impact sur la santé de façon homogène sur chaque polluant. L'agrégation avec la règle du maximum peut alors être justifiée. Les échelles de valeurs retenues pour chacun des quatre polluants entrant dans la construction de l'indice ont été établies selon les normes fixées par arrêté ministériel [2]. Ces normes permettent notamment la comparaison des différentes villes françaises grâce à un référentiel commun, même si les valeurs de l'indice dépendent aussi du choix de l'emplacement des stations permettant les relevés.

Remarquons qu'en :

- retenant comme famille de critère les quatre indicateurs qui composent l'indice Atmo,
- définissant tous les seuils de discrimination à zéro,
- n'accordant un pouvoir de veto à aucun critère,
- accordant n'importe quel jeu de poids non nuls aux critères,
- fixant le paramètre de la majorité à 1 (ce qui revient à exiger l'unanimité des votes des critères pour établir le surclassement),
- et construisant les 9 profils limites identiquement aux limites des 10 échelons de l'indice,

alors la procédure d'agrégation Max utilisée dans Atmo donne un résultat équivalent à celui de la procédure pessimiste d'ELECTRE TRI.

En effet, si les critères ne sont pas dotés de veto, la seule condition nécessaire à l'établissement du surclassement devient celle concernant l'indice de concordance : celui-ci doit alors être supérieur ou égal à 1. Or cet indice, calculé par le rapport entre la somme des poids des critères votant pour le surclassement et la somme des poids de tous les critères, est un nombre compris au sens large entre 0 et 1. Pour obtenir un indice de concordance de 1 il faut que tous les critères votent pour le surclassement. Cela revient à dire que si au moins un critère (dont le poids n'est ici pas nul) s'oppose au surclassement alors le surclassement est refusé. La procédure Max d'Atmo nous donne la pire catégorie selon tous les polluants. C'est bien cette catégorie qui sera aussi le résultat de l'agrégation d'ELECTRE TRI pessimiste car la procédure pessimiste affecte le système à la meilleure catégorie dont la limite inférieure est surclassé globalement par le système. Cette catégorie sera la meilleure catégorie pour laquelle tous les critères voteront dans le sens du surclassement de sa limite inférieure. En effet, les catégories meilleures que celle-ci ne pourront pas être acceptées car au moins un critère (celui à l'origine du Max d'Atmo) s'opposera alors au surclassement et l'indice de concordance sera alors strictement inférieur à 1.

Bien sur, une telle modélisation semble assez rigide par rapport aux possibilités offertes par ELECTRE. Ainsi, il pourrait être intéressant d'améliorer l'indice Atmo en

introduisant par exemple des seuils de discriminations non nuls ne serait-ce que pour :

- prendre en compte la mauvaise connaissance des phénomènes,
- éviter les problèmes de sauts lors de l’atteinte des valeurs limites des catégories,
- prendre en compte l’imprécision des mesures ou prévisions.

De plus, les traitements statistiques (réalisés avant l’agrégation par l’opérateur Max) permettant de conditionner les données et de résoudre d’une certaine manière le problème de l’agrégation spatiale et temporelle des différents relevés pourraient être remplacés par le principe décrit en section 3.4. Sans aller jusque là, il est pertinent de constater que nos orientations de modélisation offrent la possibilité de reproduire ce que réalise Atmo. Elles permettent de généraliser la méthode de calcul de cet indice. Finalement, ELECTRE TRI peut être considéré comme une extension naturelle d’Atmo.

Conclusion et perspectives

Nous avons tout d'abord tenté de cerner ce que peut être un processus d'aide à l'évaluation en introduisant certains concepts et en proposant une définition retenant plusieurs étapes (définir le système à évaluer, décliner un système de valeurs à travers des objectifs, sélectionner les critères mesurant l'atteinte des objectifs, évaluer les performances sur les critères, consulter les résultats de l'évaluation). Nous avons ensuite voulu montrer en quoi l'activité d'aide à l'évaluation diffère ou se rapproche de l'aide à la décision mais surtout en quoi ces deux activités peuvent se compléter. En effet, dès lors que l'on est dans un contexte où les problèmes décisionnels multicritères se succèdent fréquemment, et notamment lorsqu'on est confronté à une série de processus de décision de nature différentes mais visant à l'atteinte d'un ensemble d'objectifs communs, on peut être amené à vouloir s'appuyer sur un modèle d'aide à l'évaluation, ce qui n'exclut pas des utilisations ponctuelles de modèles d'aide à la décision. L'utilisation d'un modèle d'aide à l'évaluation s'avère alors utile à l'aide à la décision par le biais de ses fonctions (établir un diagnostic d'un système, d'alerter sur les dérives non souhaitées, suivre l'évolution du système) en permettant notamment d'identifier de nouveaux enjeux décisionnels, de faire ressortir les problèmes avec une perspective de veille ou simplement de suivre les conséquences d'une prise de décision après sa mise en œuvre. Nous pensons, plus généralement, que l'aide à l'évaluation peut être un champ de recherche très intéressant dans l'avenir car il n'a pas fait l'objet de beaucoup de travaux. De plus, et tant les domaines dans lesquels on peut pressentir des applications sont nombreux, notamment dans le suivi de politiques, dans le contrôle de gestion, dans le domaine de l'environnement non restreint aux problèmes dus aux transports, etc.

La modélisation d'un système de valeurs à travers un arbre d'objectifs est à la base de notre approche en tant qu'une des premières étapes de l'aide à l'évaluation que nous avons retenue. Cela nous semble, en effet, être un moyen particulièrement intéressant de constituer le référentiel nécessaire à l'évaluation. Voilà pourquoi, à titre exploratoire, nous avons choisi de faire un travail de terrain en appliquant une méthode d'interview. Nous avons rencontré sept personnes ayant des responsabilités publiques dans le domaine des transports et de l'environnement et qui ont accepté de jouer le jeu de la structuration hiérarchique des objectifs dans le domaine de l'environnement et des transports au niveau urbain. Le but principal était de valider la méthode de structuration du système de valeurs car cette étude avait un caractère exploratoire. Il n'y avait pas véritablement de pression au cours de ces rencontres. Les interviewés ont facilement compris la consigne de départ. Notons que le fait notamment d'appréhender le système de valeurs des interviewés en dehors de tout contexte précis (pas de projet concret, mais exploration de considérations environnementales de façon générale) a permis d'obtenir des discours qui nous ont semblés assez libre de toute contraintes, pressions et autres facteurs susceptibles d'influencer les valeurs présentées. Nous avons synthétisé ces résultats d'entretiens dans une hiérarchie d'objectifs qui a servi à illustrer la suite de notre travail. Nous avons travaillé sur des arborescences d'objectifs, et notamment réaliser des unions. Nous pensons qu'une piste de recherche peut être d'imaginer des procédures opérant sur

plusieurs arborescences d'objectifs pour par exemple les synthétiser, en faire l'union, vérifier la compatibilité de plusieurs jeux de valeurs, etc. Cela pourrait notamment aider un décideur à intégrer d'autres valeurs à son propre système.

En vue de rendre possible une agrégation multicritère, nous avons proposé d'utiliser des concepts de l'aide multicritère à la décision. Mais, il fallait légèrement les adapter à notre contexte d'aide à l'évaluation. Nous avons notamment retenu une problématique qui est un mélange des problématiques de tri et de la description, classiques en aide à la décision. Celle-ci a l'intérêt de rendre possible l'agrégation à différents niveaux tout comme la présentation des résultats désagrégés. Se posait alors le problème du choix de la méthode d'agrégation. Or, parmi les deux grandes familles de méthodes d'agrégation existantes (en aide à la décision), la plus utilisée dans le domaine de l'environnement et des transports est l'agrégation en un critère unique de synthèse. Cette famille de méthode d'agrégation n'est pas celle qui présente le plus de souplesse même si les résultats qu'elles proposent sont séduisants. Les méthodes de type ELECTRE semblent offrir plus de souplesse (présence d'incomparabilité, d'intransitivités, prise en compte de l'incertitude et de l'imprécision, limitation du phénomène de compensation) mais nécessitent en revanche de fixer un nombre important de paramètres. Elles paraissent donc plus compliquées à mettre en œuvre que les méthodes agrégeant en un critère unique, qui nécessitent simplement la construction des fonctions de transformation servant à convertir toutes les données sur une échelle unique. Cependant, il est illusoire de croire en cette simplicité si tant est que l'on souhaite faire une modélisation rigoureuse car la tâche de construction des fonctions permettant de tout ramener sur le critère de synthèse doit alors être extrêmement fine. De plus, en dépit du grand nombre de paramètres présents dans les méthodes de type ELECTRE, celles-ci permettent de simplifier la tâche de modélisation en la découpant en étapes. Notons aussi que, dans un contexte d'aide à l'évaluation ou plusieurs acteurs sont amenés à participer à la modélisation avec différents rôles (expert ou évaluant), augmenter le nombre de paramètres peut alors permettre de bien séparer les rôles de chacun à condition de bien identifier « qui doit faire quoi ». Un travail d'accompagnement de ces acteurs par des personnes averties (hommes d'études) peut alors s'avérer nécessaire. Nous avons illustré la modélisation par la construction d'un critère et des différents paramètres qui lui sont liés (catégories, seuils de discrimination et de veto) et montré en quoi ELECTRE TRI est une extension naturelle de l'indice Atmo. Le domaine de l'environnement et des transports, qui a servi à illustrer notre démarche, a beaucoup de spécificités, et en particulier il permet de soulever le problème théorique de l'agrégation temporelle et géographique de l'information. Nous avons traité ce point par une prise de position méthodologique assez simple à mettre en œuvre, mais nous sommes bien conscient que ce problème est à part entière l'objet d'un travail de recherche difficile.

Références

- [1] Abernot Y., 1996, Les méthodes d'évaluation scolaire, Dunod, Paris.
- [2] ADEME, 2000, Atmo 2000, Directive de 10 janvier 2000, Ministère de l'environnement, Paris.
- [3] Berge C., 1970, Graphes et hypergraphes, Dunod Université N°604.
- [4] Bouyssou, D., Marchant, T., Pirlot, M., Perny, P., Tsoukiàs, A., Vincke, Ph. (2000). Evaluation and decision models : a critical perspective. Dordrecht : Kluwer Academic.
- [5] Bouyssou, D., Marchant, T., Pirlot, M., Tsoukiàs, A., Vincke, Ph. (2007). Evaluation and decision models : stepping stones for the analyst. Dordrecht : Kluwer Academic.
- [6] Cardinet J., 1992, L'objectivité de l'évaluation, Formation et technologies \acute{U} Revue européenne des professionnels de la formation, n°0, p17-26.
- [7] Chakhar S. 2006, Cartographie Décisionnelle Multicritère : Formalisation et Implémentation Informatique, thèse de doctorat, Lamsade, Université Paris Dauphine.
- [8] Chovin, P., Richalet, J., 1973. Étude théorique de la cinétique de la fixation du monoxyde de carbone sur l'hémoglobine du sang. Ann. Fals. exp. chim., 710, 177-189.
- [9] Habermas J., 1987, Théorie de l'agir communicationnel, Fayard, Paris.
- [10] De Ketele J.M., Gérard F.M., Roegiers X., 1997, L'évaluation et l'observation scolaire : deux démarches complémentaires, Education \acute{U} Revue de diffusion des savoirs en éducation, n°12, p33-37.
- [11] Garcia J., Colosio J., Jamet P., 2001, Les indices de la qualité de l'air, Ecole des Mines de Paris, Paris.
- [12] Gastaut G., Maugard A., Agard J. al, 1970, Rationalisation des choix budgétaires, Monographies de Recherche Opérationnelle, Dunod, Paris.
- [13] Gérard J.M., 2002, L'indispensable subjectivité de l'évaluation, Antipodes, n°156, p26-34.
- [14] Keeney R.L., 1992, Value Focused Thinking - A Path to Creative Decisionmaking, Harvard University Press, Cambridge.
- [15] Maurin, M., 1981. Modélisation de la réponse physiologique à l'exposition au monoxyde de carbone. Rapport IRT-CERNE, NNP 66.
- [16] Maurin M., 2000, L'agrégation des indicateurs de l'environnement, quelques premières réflexions, INRETS, Bron.
- [17] Maurin, M., 2007. Les indicateurs d'impact sur l'environnement, compléments sur les paniers et sur les lois d'impact. Note INRETS-LTE, Bron.
- [18] Mazri C., 2007, Apports méthodologiques pour la construction de processus de décision publique en contexte participatif, Lamsade, Université Paris Dauphine, Paris.
- [19] Moussa N., 2001, Aide multicritère à l'évaluation qualitative par inférence de modèles de tri ordinal sur une hiérarchie de critères \acute{U} Application pour la conception

d'un système européen de cotation de PME technologiques innovantes, thèse, Lam-sade, Université Paris Dauphine, Paris.

- [20] Noizet G., Caverny J.P., 1978, La psychologie de l'évaluation scolaire, Presse Universitaire Française, Paris.
- [21] Pujol P., 2000, Prospective et Indicateurs d'impact des transports sur l'Environnement : outils d'évaluation et d'aide à la décision - Enquête utilisateur, Inrets, Bron.
- [22] Rousval B., 2005, Aide multicritère à l'évaluation de l'impact des transports sur l'environnement, Thèse de doctorat, Université Paris Dauphine, Paris.
- [23] Rousval B., Maurin M., 2008, Evaluation de l'impact des transports sur l'environnement, quels modèles utiliser ?, revue RTS n°100, Inrets, Bron.
- [24] Roy B., 2006, Communication orale, Conf. Les outils pour décider ensemble : nouveaux territoires, nouveaux paradigmes, Université Paris Dauphine, Paris.
- [25] Roy B., Bouyssou D., 1993, Aide multicritère à la décision : méthodes et cas, Economica, Paris.
- [26] Roy B., Figueira J, 2001, Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure, European Journal of Operational Research, n°139, p317-326.
- [27] US-DoT, 1975. A guidance manual for the assessment of physical impacts due to highway facility improvements. US Department of Transportation, Washington DC, États-Unis.
- [28] Vial M., 1999, Modèle et logique de l'évaluation, colloque international, Ethique et qualité dans l'évaluation, 25-26 octobre, Université de Reims Pol'èvalue, Reims.

Annexes : principes d'ELECTRE TRI

Nous présentons ici certains principes de la méthode ELECTRE TRI proposée par Roy et Bouyssou[25]. Cette méthode consiste à réaliser une agrégation multicritère en affectant une action à une des catégories qu'on aura pris soin de définir préalablement. Elle permet de répondre ainsi à la problématique de tri.

Considérons tout d'abord, les trois relations binaires suivantes, permettant une comparaison fine de deux actions a et b , localement à un critère g :

- a est indifférente à b (et inversement) : relation (aI_gb) ,
- a est préférée strictement à b (ou inversement) : relation (aP_gb) ,
- a est préférée faiblement à b (ou inversement) : relation (aQ_gb) .

Pour traduire ces cas de figures, considérons alors un premier seuil sur l'échelle du critère g . Le seuil d'indifférence q tel que, lorsque $g(a) > g(b)$:

- $(aP_gb) \iff g(a) - g(b) > q$,
- $(aI_gb) \iff g(a) - g(b) \leq q$.

Ce seuil représente l'écart à partir duquel on commence à préférer l'une des deux actions. Si la performance de a est la même que celle de b ou si a est meilleur que b de moins de la quantité q alors on est indifférent entre a et b . Le seuil q doit être fixé de sorte que si l'action a est meilleure que b de plus de la quantité q alors on a une préférence stricte de a sur b . Un critère g muni d'un tel seuil est alors appelé un quasi-critère.

Pour faire ensuite apparaître la situation de *préférence faible* pour permettre de traduire, dans la modélisation d'un critère, l'hésitation du décideur entre la situation d'*indifférence* et celle de la *préférence stricte* et ainsi, affaiblir encore le pouvoir discriminant du critère, considérons un second seuil : le seuil de préférence p . On a donc pour $g(a) > g(b)$:

- $(aP_gb) \iff g(a) - g(b) > p$,
- $(aQ_gb) \iff q < g(a) - g(b) \leq p$,
- $(aI_gb) \iff g(a) - g(b) \leq q$.

L'introduction conjointe des seuils d'indifférence et de préférence fait de g un *pseudo-critère*.

L'agrégation multicritère est fondée sur la comparaison par paire d'action a et b . Pour cela, on confère à chacun des critères de la famille F , deux pouvoirs :

- un pouvoir de vote en faveur du surclassement global de b par a : le pouvoir de vote noté k_i est fixé pour chaque critère g_i . La voix ainsi exprimée par le critère sera pondérée par un poids qui représentera l'importance relative que les décideurs apportent à ce critère par rapports à tous les autres de la famille F .
- un droit de veto contre le surclassement : sur un critère, le seuil de veto est la valeur telle que si l'action a est éloignée défavorablement de plus d'une fois ce seuil de la performance de l'action b , alors on refuse le surclassement (aSb) et cela quelles que soient les performances sur les autres critères. Le pouvoir de veto n'est pas forcément conféré à tous les critères, toujours selon les préférences des décideurs.

Lors de la comparaison globale (ou multicritère) de deux actions a et b , les méthodes de type Electre s'intéressent à la question du surclassement (aSb), c'est-à-dire à la question « l'action a est-elle globalement au moins aussi bonne que l'action b ? ».

Pour cela, cette question est tout d'abord posée à chacun des critères (pseudocritères), qui va voter pour ou contre cette assertion. Considérons alors la relation de *surclassement* S_g entre deux actions sur un pseudo-critère g définie de la façon suivante :

$$- (aS_gb) \iff (aI_gb) \text{ ou } (aP_gb) \text{ ou } (aQ_gb)$$

Cette relation traduit le fait que sur le critère g , l'action a est *au moins aussi bonne que* l'action b . On dit alors que a surclasse b localement au critère g . Remarquons que si $g(a)$ est juste au dessous de $g(b) + q$ alors $(bS_g a)$ est établi mais si $g(a)$ est juste au dessus de $g(b) + q$ alors $(bS_g a)$ n'est plus vérifiée. Pour éviter ce saut au niveau du seuil d'indifférence dans la comparaison $(bS_g a)$ regardons le nombre flou φ défini comme suit :

- φ vaut 1 quand $(bI_g a)$ ou $(bQ_g a)$ ou $(bP_g a)$,
- φ vaut 0 quand $(aP_g b)$,
- φ décroît linéairement de 1 à 0 quand $(aQ_g b)$.

On notera cette fonction $\varphi(bS_g a)$. Elle est à valeurs dans $[0; 1]$ et représente le degré avec lequel l'assertion b surclasse a est vraie selon le critère g . On peut de même, construire $\varphi(aS_g b)$ qui représentera le degré avec lequel l'assertion a surclasse b est vraie toujours selon le critère g .

On considèrera finalement la relation (aSb) comme vérifiée globalement si une *majorité* de critères a voté en ce sens.

La seconde condition pour accepter le surclassement global (aSb) est qu'aucun des critères parmi ceux n'ayant pas voté pour l'établissement de (aSb) ne s'oppose trop fortement à la majorité en apportant son veto. Autrement dit, le décideur peut, dans la modélisation, conférer un pouvoir de veto à certains critères. Le sens du veto est : le critère g s'oppose à laisser dire à une majorité d'autres critères que l'action a est au moins aussi bon que b , si, sur l'échelle de g , l'action b est vraiment meilleure que l'action a . Pour refléter cela on doit, pour chaque critère g_i , déterminer le seuil de veto v_i , c'est-à-dire la valeur sur le critère à partir de laquelle on considère que l'écart entre a et b est tellement important qu'il conduit à refuser l'assertion de surclassement globale de la meilleure action par la moins bonne.

Pour la comparaison globale de l'action b avec l'action a , chaque critère g_i de la famille F est finalement muni de :

- la fonction $\varphi_i(bS_{g_i} a)$ à valeurs dans $[0; 1]$,
- un poids k_i reflétant son pouvoir de vote,
- un éventuel seuil v_i lui conférant un pouvoir de veto au surclassement global.

Pour prendre en compte les votes des critères et les éventuels veto, considérons maintenant les deux indices suivants à valeurs dans l'intervalle $[0; 1]$:

- l'indice de concordance de la famille de critère F noté $c(b, a)$,
- l'indice de discordance d'un critère g_i noté $d_i(b, a)$.

L'indice de concordance peut être défini ainsi :

$$c(b, a) = \frac{\sum_{i \in F} \varphi_i(bS_{g_i}a) \times k_i}{\sum_{i \in F} k_i}$$

On obtient donc par ce rapport un nombre compris en 0 et 1 représentant la coalition de critères votant pour le surclassement.

L'indice de discordance a pour définition formelle :

- $d_i(b, a) = 1$ si $g_i(a) > g_i(b) + v_i$,
- $d_i(b, a) = 0$ si $g_i(a) \leq g_i(b) + p_i$,
- $d_i(b, a) = \frac{g_i(a) - g_i(b) - p_i}{v_i - p_i}$ sinon.

Il représente le degré avec lequel un critère g_i s'oppose à l'établissement du surclassement global (bSa). Cet indice prend la valeur 1 quand l'action a est meilleure que l'action b de plus de la valeur du seuil de veto v_i . Il prend la valeur 0 quand b est au moins aussi bon que a sur le critère g_i c'est-à-dire quand ($bS_{g_i}a$) est établi (cas où l'action b est soit préférée, même faiblement, soit indifférente à l'action a). Entre 0 et 1, l'indice de discordance croît proportionnellement à $g_i(a) - g_i(b)$.

Alors, pour la comparaison de deux actions a et b , la proposition (bSa) est vérifiée si et seulement si la condition suivante est réalisée : $\sigma(a, b) \geq \lambda$ avec $0 \leq \lambda \leq 1$.

La fonction σ étant l'indice de crédibilité de la proposition (aSb) utilisé dans Electre III et défini par :

- $\sigma(a, b) = c(b, a) \times \prod_{i \in D} \frac{1 - d_i(b, a)}{1 - c(b, a)}$ si $D = \{i \in F : d_i(b, a) > c(b, a)\} \neq \emptyset$
- $\sigma(a, b) = c(b, a)$ si $D = \emptyset$

Ainsi, une première condition nécessaire au surclassement est que le pourcentage des votes allant pour l'établissement du surclassement (pourcentage représenté par l'indice de concordance $c(b, a)$) soit supérieur ou égal à λ . Dans Electre, le nombre λ est un paramètre du modèle compris aussi entre 0 et 1. Fixer par exemple le paramètre λ à 0,7 revient à exiger au moins 70% des votes qui soient favorables au surclassement (votes des critères de la famille F , pondérations prises en compte). Ce paramètre représente finalement la sévérité que le décideur impose au modèle pour fournir un résultat agrégé. Plus λ est proche de 1 plus la sévérité est grande. La seconde condition est que le terme produit multiplié à $c(b, a)$ ne diminue pas la valeur de σ en dessous de la valeur de λ , c'est-à-dire qu'aucun critère ne s'oppose trop fortement à l'établissement du surclassement (bSa).

Etablir que (bSa) revient à conclure que b est au moins aussi bon que a . Quand les surclassements (bSa) et (aSb) sont établis simultanément alors on conclut à l'indifférence (aIb). En revanche, quand ni (bSa) ni (aSb) ne sont vérifiées alors on conclut à (aRb) : c'est le cas de l'incompatibilité. Dans les autres cas, l'une des deux actions (et seulement l'une des deux) surclasse l'autre : c'est la préférence. On peut finalement résumer les situations de comparaison globale entre deux actions a et b à ces quatre cas :

- quand $\text{Non}(aSb)$ et $\text{Non}(bSa)$, incomparabilité : (aRb) \iff (bRa),

- quand (aSb) et $\text{Non}(bSa)$, préférence pour a : (aPb) ,
- quand $\text{Non}(aSb)$ et (bSa) , préférence pour b : (bPa) ,
- quand (aSb) et (bSa) , indifférence : $(aIb) \iff (bIa)$.

Répondre à la problématique du tri, c'est affecter chaque action dans une des catégories ordonnées préalablement définies. ELECTRE TRI est une méthode de type ELECTRE développée pour répondre à la problématique du tri, c'est-à-dire affecter chaque action à une des catégories ordonnées prédéfinies. Ce découpage doit se faire sur les échelles de tous les critères de la famille F .

ELECTRE TRI nécessite de bâtir un ensemble de catégories ordonnées à l'aide de profils limites. Les profils limites sont des actions fictives et bien choisies, délimitant chacune des catégories. Quatre actions fictives peuvent ainsi délimiter cinq catégories ordonnées comme par exemple, de la meilleure à la moins bonne :

- ensemble des actions très intéressantes au sens de la problématique,
- ensemble des actions assez intéressantes au sens de la problématique,
- ensemble des actions moyennement intéressantes au sens de la problématique,
- ensemble des actions peu intéressantes au sens de la problématique,
- ensemble des actions très peu intéressantes au sens de la problématique.

Pour cela, il s'agit de fixer les quatre valeurs limites sur chaque critère. Ensuite, il est possible d'affecter chacune des actions réelles à une des catégories prédéfinies grâce à la comparaison aux actions fictives, profils limites.

L'affectation de l'action à une catégorie réalisée dans ELECTRE TRI nécessite la réalisation préalable des étapes suivantes :

- construire une famille F de critère,
- définir les seuils (de préférence, d'indifférence et de veto),
- pondérer les critères,
- fixer le paramètre λ , majorité requise pour établir le surclassement,
- construire de n profils limites définissant $n + 1$ catégories.

La procédure d'affectation à une catégorie utilise alors la relation binaire de surclassement global S définie précédemment servant à la comparaison de deux actions a et b . Etant donné une action a , ELECTRE TRI se propose alors de la comparer à l'ensemble des profils limites, qui sont des actions fictives séparant les catégories. Deux procédures sont alors possibles pour affecter une action a à une des $n + 1$ catégories et qui utilisent la condition nécessaire au surclassement (concernant l'indice de crédibilité et le paramètre λ). La procédure pessimiste affecte alors l'action à la meilleure catégorie dont le profil limite inférieur est surclassé globalement par a . La procédure optimiste affecte l'action à la moins bonne catégorie dont le profil limite supérieur surclasse globalement a et n'est pas surclassé globalement par a . Il en résulte que l'affectation optimiste est au moins aussi bonne que l'affectation pessimiste.