



Université Paul Sabatier
IRIT-LILaC
Andreas Herzig
118 route de Narbonne
31062 Toulouse Cedex 9

Tel. : 05 61 55 81 23
Fax : 05 61 55 88 98
Mél : herzig@irit.fr

Rapport sur la thèse de Jonathan BEN-NAIM « Sur la caractérisation du raisonnement du sens commun en présence d'information incertaine »

La thèse de Jonathan BEN-NAIM traite de deux sujets centraux en intelligence artificielle, à savoir du raisonnement non-monotone et de la révision. Ces deux sujets constituent la première et la deuxième partie du manuscrit.

En ce qui concerne le raisonnement non-monotone, les premiers travaux de Reiter et col. au début des années 80 avaient pour objectif la définition d'opérations d'inférence concrètes permettant de raisonner par défaut. Plus tard, Gabbay, Lehmann, Makinson et col. ont posé la question des propriétés générales de telles opérations. Ils y ont donné une première réponse en les basant sur des familles de sémantiques abstraites (en définissant en particulier les systèmes d'inférence préférentiels et rationnels). Ces sémantiques reposent sur celle de la logique classique propositionnelle. La première partie de thèse de Jonathan BEN-NAIM prolonge ces approches, en les généralisant vers des logiques à plusieurs valeurs de vérité. Tandis qu'une telle généralisation a déjà été proposée par des auteurs de la littérature récente (qui sont dûment cités), la présente thèse va au-delà, en posant un cadre général pour une telle entreprise, et en fournissant des résultats généraux de caractérisation de classes de structures par des ensembles de propriétés. Ainsi, le 3^{ème} chapitre contient une caractérisation des relations d'inférence basées sur les préférences, et le 4^{ème} chapitre contient une caractérisation des relations d'inférence à pivot récemment introduites par Makinson. Tous ces résultats sont originaux, et ils sont rigoureusement démontrés. Leur valeur est également attestée par le fait que le contenu des chapitres 3 et 4 a été publié dans le *Journal of Logic and Computation*.

En ce qui concerne la révision des croyances, mis à part les opérations de révision basées sur la syntaxe (auxquels il est difficile d'associer une sémantique), on peut dire que l'opération concrète la plus connue est celle dite de Dalal basée sur la distance de Hamming de 1988. Sur le plan des propriétés abstraites, les premiers travaux de Alchourron, Gardenfors et Makinson (1985) posaient une sémantique en termes de préférences (alias « enracinement épistémique ») pour de telles opérations, et les caractérisaient par un ensemble de postulats dits « postulats AGM ». Cette sémantique a été critiquée comme étant trop libérale, en particulier parce qu'elle contraint très peu l'itération de révisions. Ce n'est qu'avec les travaux de Schlechta et col. (1996) qu'une sémantique basée sur les distances généralisant celles de Dalal a été systématiquement étudiée. Leur résultat de caractérisation contenait un postulat infinitaire appelé *loop*, et il était tout naturel de poser la question si une

caractérisation finitaire est possible. De telles questions sont faciles à poser et souvent extrêmement difficiles à répondre. C'est le cas ici. Après une première réponse négative de Schlechta en 2004, la thèse de Jonathan BEN-NAIM contribue à une réponse négative très générale : sommairement, il démontre qu'il ne peut exister une relation à arité finie caractérisant la famille des opérations de révision basées sur une distance. Dus à leur nature négative, de tels résultats très techniques sont souvent plus difficilement acceptés dans des conférences ; le fait que l'article décrivant ce résultat a été accepté dans la conférence *Knowledge Representation and Reasoning (KR'2006)* souligne sa qualité et son importance.

Le manuscrit de la thèse est d'une qualité exemplaire. Dans chacun des chapitres la littérature pertinente est citée et discutée d'une manière complète. Les définitions et théorèmes sont introduits avec clarté, précision et concision. Les démonstrations des théorèmes mettent en évidence que Jonathan BEN-NAIM maîtrise parfaitement les outils formels dont il se sert dans son travail. Toutes ces qualités traduisent une excellente maîtrise de l'étendu du sujet traité, et une grande culture dans la littérature du domaine.

Le seul point qu'on pourrait soulever concerne le lien entre les deux parties de la thèse. En effet, elles apparaissent complètement séparées, autant sur le plan des outils formels (logiques à plusieurs valeurs dans la partie I vs. logique classique dans la partie II; préférences en I, distances en II). La conclusion mentionne bien que les résultats de la partie II devraient être généralisables aux logiques à plusieurs valeurs, mais il me semble que bien plus pourrait être dit. En effet, depuis l'article de Makinson et Gardenfors au *Nonmonotonic Reasoning Workshop* de 1988 il est connu qu'il existe un lien étroit entre inférence non-monotone et révision : en identifiant « B est conséquence de A » avec « la révision de C par A contient B » (où C est une théorie dite d'arrière-plan) on peut établir une correspondance systématique entre postulats pour l'inférence non-monotone et postulats pour la révision. Une telle 'passerelle' devrait par exemple permettre d'exporter la sémantique à plusieurs valeurs vers la révision (piste mentionnée dans la conclusion) d'une manière automatique. D'autres mises en correspondance pourraient d'ailleurs également constituer des voies de recherche intéressantes. Un premier candidat serait les logiques des conditionnels, dont les axiomes entretiennent des liens étroits avec les postulats pour l'inférence non-monotone (cf. l'article de D. Makinson, « Five faces of minimality »). Un second serait la mise à jour : en effet, tout comme les opérations de révision AGM, la sémantique générale pour les opérations de mise à jour dite KM (Katsuno et Mendelzon, KR'92) est basée sur les préférences, tandis que les opérations concrètes comme celle dite de Winslett (AAAI'88) et celle dite de Forbus (AAAI'88) sont toutes les deux basées sur la distance de Hamming. Il serait à mon avis intéressant de se poser la question si les résultats d'impossibilité de la partie II de la thèse ont une contrepartie ici.

En résumé, Jonathan BEN-NAIM réussit son entreprise d'une manière convaincante, en faisant preuve d'une grande technicité et d'une finesse remarquable dans la manipulation d'outils formels très complexes. Il s'agit clairement d'une thèse exceptionnelle : c'est une des meilleures que j'ai vu dans les 7 dernières années. Je recommande sans réserves de lui attribuer le degré de docteur de l'université en informatique.

Toulouse, le 3 avril 2006



Andreas Herzig
Directeur de recherche CNRS