

## Proposition de thèse

**TITRE DE LA THESE**

Modèle génératif cognitif de l'argumentation  
Application aux agents conversationnels et à l'aide contextuelle

**RESPONSABLE(S) DE LA THESE**

[Jean-Louis Dessalles](#)

**DOCTORANT :****EQUIPE(S) D'ACCUEIL OU SE DEROULERA LA THESE**

Groupe Interaction, Cognition et Complexité (IC2) – Dép. InfRes, TELECOM ParisTech

**PROJET (S) STRUCTURANT (S) DE RATTACHEMENT**

DBWeb (base de données et Web)

**MOTS CLES**

Argumentation, Agents conversationnels, Aide contextuelle, Traitement du langage naturel, Cognition, Pertinence

**RESUME :**

L'argumentation occupe plus de la moitié du temps de parole des êtres humains. L'une des tâches essentielles de l'intelligence artificielle est d'amener les machines à argumenter correctement. Or, les techniques actuelles ne cherchent pas à imiter le mode argumentatif humain.

L'objectif du traitement automatique de l'argumentation est d'amener la machine à analyser et à produire des arguments pertinents *pour l'utilisateur*. Les techniques actuelles de traitement de l'argumentation reposent pour l'essentiel sur le traitement de dialogues finalisés (orientés vers la résolution d'une tâche spécifiée à l'avance). Cela revient à produire des arguments pertinents *du point de vue de la machine*. Une telle limitation conduit les concepteurs d'agents conversationnels à utiliser de simples techniques d'appariement, en se contentant de sélectionner les interventions de la machine dans une liste d'arguments possibles prévus à l'avance.

L'objectif de la thèse est de développer un *modèle génératif des arguments pertinents* qui permettra un authentique calcul d'arguments *cognitivement* pertinents en fonction du contexte. Les éléments théoriques d'un tel modèle sont déjà disponibles. Il s'agit, dans le cadre de cette thèse, de les améliorer puis de les implanter en situation réelle (agents conversationnels et systèmes d'aide contextuels).

L'enjeu direct est d'améliorer significativement la performance des agents conversationnels tels qu'ils sont de plus en plus couramment utilisés sur la Toile. L'enjeu indirect est de fournir une couche supplémentaire au Traitement du Langage Naturel, utilisable pour améliorer la compréhension, la traduction ou le résumé.

## DESCRIPTIF DETAILLE

### ◆ Contexte Scientifique

Le dialogue homme-machine connaît depuis quelques années un renouvellement dû au développement d'une technique de dialogue empirique : les agents conversationnels (*chatbots*). Ignorant les techniques récentes et reprenant à leur compte des travaux historiques comme les solutions de filtrage de type *Eliza*, ces systèmes sont capables de performances assez surprenantes, pour autant que la tâche de dialogue ne soit pas trop difficile. Ces approches ont même donné lieu à un certain nombre d'applications commerciales. Il existe également une compétition/évaluation de ce type de systèmes, qui se tient régulièrement (sur le modèle de compétitions plus scientifiques de type TREC, MUC, aux USA) : le *Loebner Prize*.

Les approches de type *chatbot* souffrent cependant de limitations évidentes, en particulier en ce qui concerne les modèles de dialogue adaptables à des dialogues ouverts, par exemple les systèmes d'aide ou les systèmes de conseil à la vente, pour lesquels la simple illusion d'une conversation ne suffit pas. Dès que l'utilisateur est en droit d'attendre du dialogue qu'il soit utile, il faut s'assurer que les interventions du système soient *pertinentes*.

La production d'arguments pertinents ne peut être le fruit d'une opération imprécise, qui se fonderait par exemple sur l'imitation statistique. Au contraire, rien n'est laissé au hasard lors d'une prise de parole argumentative, car l'acceptation sociale de son auteur en dépend. L'individu conçoit son intervention pour qu'elle s'ancre logiquement sur les interventions précédentes. Il est donc naturel de postuler l'existence d'une procédure cognitive particulière qui rend les humains capables de déterminer précisément le contenu de leurs interventions langagières.

Les approches classiques reposent sur des hypothèses particulières, trop fortes pour de nombreuses applications. Par exemple, il est généralement supposé que le système doit avoir accès à un ensemble pré-calculé de plans (Bratman *et al.* 1988), ce qui suppose une combinatoire qui n'explose pas. De plus, il doit être possible d'effectuer des opérations globales sur ces ensembles de plans, comme des tests de cohérence entre plans (Amgoud *et al.* 2003). Le système utilise généralement un algorithme de planification, par exemple un analyseur moyens-fins, capable de produire une série de plans qui sont ensuite filtrés selon leur compatibilité avec les intentions du moment (Bratman *et al.* 1988). On suppose souvent que le système a accès à des relations statiques entre arguments et qu'il peut effectuer des opérations globales sur le graphe de ces relations, comme découvrir des circuits ou des chemins indépendants (Dung 1995). Le système est parfois supposé utiliser des ensembles de patrons d'arguments ou des ensembles de règles épistémiques et comportementales appelés jeux de dialogue (Airenti *et al.* 1993; Hulstijn 2000; Maudet 2001). Les systèmes d'argumentation utilisent également des règles complexes leur permettant de prêter des croyances aux autres participants du dialogue (Lee & Wilks 1996). Ils font aussi appel à une expertise sociale qui leur permet de manipuler des « engagements sociaux » (Bentahar *et al.* 2003). Dans certains systèmes, l'expertise sociale n'est pas séparée formellement de l'expertise du domaine et ce sont les mêmes procédures qui sont supposées agir sur les deux. Certains systèmes, enfin, utilisent une référence présentée comme objective (oracle), ce qui leur permet de définir une notion de vérité et d'utiliser la sémantique de la théorie des modèles pour évaluer les croyances des participants (Shapiro *et al.* 2000) et leur progression vers la « vérité » et le consensus (Tallon *et al.* 2003).

Certaines de ces hypothèses peuvent être satisfaites dans des situations très contraintes, comme les dialogues finalisés. Dans ce cas, les contraintes importées de la tâche à résoudre sont suffisantes pour que les interventions dialogiques soient possibles. En revanche, la plupart de ces suppositions se révèlent irréalistes lorsqu'il s'agit de concevoir un agent conversationnel dont la tâche est, par exemple, d'aider un utilisateur sur un site qu'il ne connaît pas ou dans un système d'aide contextuel. Le système ne peut pas partir de son propre problème (la tâche à accomplir), mais doit comprendre le problème de l'utilisateur avant de s'attacher à proposer une solution.

Le défi d'un modèle génératif de l'argumentation est d'expliquer comment l'on peut passer d'une connaissance statique, prêtée aux interlocuteurs, à une argumentation co-construite dans laquelle chaque argument est pertinent. La "règle du jeu" de ce type de modélisation est assez stricte. Elle peut être déclinée ainsi.

– Dialogue non finalisé : L'interaction n'est pas routinière. Contrairement à ce qui prévaut pour des situations cadrées comme l'achat d'un billet de train au guichet, les participants ne disposent pas de script pour savoir comment se comporter. Ils doivent donc, à proprement parler, calculer leurs interventions.

– Localité : La connaissance à partir de laquelle ce calcul s'opère ne peut être *a priori* circonscrite. Idéalement, elle est produite à la demande à partir de modèles plus fins (simulations qualitatives ou analogiques). Si, par commodité, cette connaissance est stockée sous la forme d'un sur-ensemble de règles, on impose qu'il

soit adressé par le contenu et ne puisse pas être parcouru séquentiellement. On s'interdit, en particulier, d'effectuer des opérations globales comme des tests de cohérence ou des recherches de cycle.

– Problème vs. buts : La même restriction s'applique aux croyances et aux désirs. Les participants n'abordent pas une discussion avec une interminable liste de buts explicites comme rester en vie, ne pas perdre sa carte de crédit, etc. Les croyances et les désirs pertinents sont activés au moment de l'interaction, et les buts sont engendrés en réponse à des *problèmes*.

– Simplifications : La connaissance du domaine est fournie à la procédure argumentative à la demande, quoique seulement à travers un adressage par le contenu. Par ailleurs, dans un premier temps, les aspects métacognitifs peuvent être ignorés, ce qui revient à considérer que la connaissance du domaine est partagée par les participants et que les calculs épistémiques sont évités. La procédure est de plus considérée comme *autonome*, c'est-à-dire qu'elle est indépendante d'autres aspects de l'interaction comme le respect de la face (Muntigl & Turnbull 1998), les attitudes coopératives, la politesse ou plus généralement la gestion des relations sociales.

Dans l'approche qu'il est proposé de suivre dans cette thèse, la génération d'argument tire ses contraintes de principes cognitifs en non du seul cadre limitatif d'une tâche préétablie à résoudre. Elle se fonde sur la notion de *conflit cognitif* (contradiction entre croyances et/ou désirs). La gestion de l'argumentation oscille entre la détection des conflits cognitifs et leur solution.

#### ◆ Contenu Scientifique

Nous avons commencé la modélisation de l'argumentation dès les années 1990. Nous avons proposé un système argumentatif pour l'aide à l'apprentissage (Dessalles 1990a) et un premier modèle de génération d'arguments pertinents (Dessalles 1990b). L'étude approfondie d'un corpus de dialogues spontanés humains nous a permis de dégager un modèle minimaliste de la génération d'argument. Le modèle auquel nous sommes parvenus consiste en une procédure comportant trois phases : *détection de conflit cognitif, abduction, négation* (Dessalles 1998, 2001, 2008).

Un *conflit cognitif* est constaté dans une situation où des affirmations successives se contredisent, c'est-à-dire si une proposition et sa négation se voient attribuer des affirmations de même signe. Le conflit a une *intensité*, donnée par l'affirmation de son *terme faible*. L'affirmation d'un terme est calculée par contraste entre la croyance/désir du terme et la croyance/désir de sa négation. La croyance se mesure par l'opposé de l'inattendu (Dessalles 2006).

L'*abduction*, comprise comme production de cause, est une phase essentielle de la procédure argumentative. L'abduction est connue comme constituant un aspect central de la résolution de problème (Magnani 2001), du raisonnement diagnostique (Reiter 1987) et plus généralement de l'intelligence humaine (Hobbs, Stickel & Martin 1993 ; Hobbs 2004). La notion d'abduction est prise ici dans une acception très proche de celle que lui a donnée Charles Pierce : génération d'une hypothèse explicative. Noter que l'acte abductif lui-même est extérieur au modèle, et différentes techniques d'abduction peuvent être utilisées.

La *négation* consiste à renverser la perspective de résolution : si la remontée aux causes d'un problème (phase diagnostique) se révèle infructueuse, le renversement de perspective consiste à rechercher de possibles causes pour produire la négation du problème (phase créative).

La procédure de génération d'argument minimale se déroule comme suit : (1) détection d'un conflit cognitif ; (2) abduction à partir du terme faible et propagation du conflit sur la cause la plus faible si elle existe ; (3) négation et retour en phase 2. La sortie de la procédure s'effectue lorsque la propagation permet de faire chuter l'intensité du conflit.

La procédure produit une solution (explication ou suggestion) lorsque la propagation du conflit tombe sur une cause dont l'affirmation est de même signe que l'affirmation de son effet. Lorsque ce n'est pas le cas, la procédure continue à rechercher des causes qui font baisser l'intensité du conflit.

Comme l'argumentation humaine, la procédure est susceptible de « boucler » de deux manières. L'une vient de ce qu'une solution au conflit courant risque d'engendrer elle-même un conflit. C'est souvent le cas dans les problèmes contraints. Ce premier bouclage est la source principale de la richesse de l'argumentation. L'autre source de bouclage, moins féconde, apparaît si on l'autorise le système à réévaluer les affirmations qui résistent à l'abduction. La sortie de boucle est alors rendue possible par une élévation graduelle du seuil de conflit.

Cette procédure Conflit–Abduction–Négation a déjà donné lieu à une implémentation de type « laboratoire » (Dessalles 2008). Ce programme est capable, pourvu qu'on lui fournisse une connaissance sous forme symbolique (règles causales), de produire les mêmes arguments que ceux qui sont observés dans des dialogues réels, en utilisant un nombre réduit d'étapes. Cette performance ne doit pas être sous-estimée. Généralement, les procédures de planification utilisées pour l'argumentation considèrent successivement de nombreuses possibilités inutiles et, contrairement aux êtres humains, fondent leurs choix sur en évaluant simultanément de multiples options. Le défi auquel s'attaque le type d'approche proposé ici n'est pas seulement de produire une argumentation correcte, mais également de la produire en un nombre d'étapes minimal au moyen d'une procédure de complexité minimale.

L'un des atouts majeurs de la procédure Conflit–Abduction–Négation est sa plausibilité cognitive. Celle-ci est essentielle d'une part pour permettre l'analyse des arguments de l'utilisateur et d'autre part pour la production d'arguments humainement acceptables, autrement dit pertinents.

Le principal point fort de l'approche Conflit–Abduction–Négation, outre sa plausibilité cognitive, est que toutes les décisions prises par la procédure argumentative sont *séquentielles* et *locales*. À chaque étape, la procédure « tient » un prédicat et prend une décision du type « chercher une cause », « nier », « proposer une solution » ou « réviser l'affirmation ». Il n'y a pas d'optimisation globale. De ce fait, le modèle échappe au problème du cadre (Chiappe & Kukla 1996) et au problème de la complexité de la satisfaction de contraintes.

Le travail envisagé dans le cadre de cette thèse comporte une partie théorique et une partie pratique. Il s'agit, sur le plan théorique, de tester, valider et compléter la procédure Conflit–Abduction–Négation dans des situations de dialogue ouvert, dans lesquels la machine dispose d'une connaissance experte. Une part de cette réflexion concernera la distinction entre les interventions qui doivent être rendues explicites et celles qui doivent être tues, de manière à éviter au programme d'un côté le risque d'être « lourd » et de l'autre côté le risque d'être elliptique (actuellement, les principes qui gouvernent la décision d'explicitation font défaut). Cette phase pourra être testée dans la reconstruction de dialogues réels et ainsi que dans l'interactivité.

La partie pratique de la thèse portera sur la réalisation d'un système d'aide contextuel sous la forme d'un agent conversationnel. Il ne s'agit pas seulement de réaliser un agent particulier, mais un agent générique capable d'argumenter à propos de toute connaissance experte qui lui sera fournie en entrée. Il faudra bien entendu prévoir les outils de traitement du langage naturel nécessaires à l'interaction. Ce traitement devra permettre une certaine souplesse d'expression, qui est possible dans les domaines dont la sémantique peut être assez bien représentée par un système de règles (ce qui inclut la possibilité d'ontologies).

L'originalité de l'approche, sur le plan technique, est liée à la séparation stricte entre la connaissance argumentative, qui reste fixe et incluse pour l'essentiel dans la procédure d'argumentation, et la connaissance du domaine qui, elle, varie totalement d'une application à l'autre. C'est en ce sens que l'agent conversationnel projeté est générique. Cette généralité est rendue possible par la conception préalable d'un modèle argumentatif général indépendant du contenu.

L'argumentation de l'agent conversationnel est prévue pour apporter une aide contextuelle. La situation est propice à l'argumentation, car l'utilisateur démarre l'interaction avec un problème à résoudre. La première tâche de l'agent est d'identifier le problème, et de poser des questions s'il n'est pas apparent dès la première requête ou s'il est ambigu. Ensuite, la procédure conflit–abduction–négation tente d'apporter une solution par l'interaction. L'apport principal de ce système d'aide est qu'il est supposé comprendre pourquoi ses premières suggestions sont elles-mêmes problématiques, ce qui est généralement le cas lorsque le problème de l'utilisateur n'est pas trivial. Grâce à la procédure argumentative, le système et l'utilisateur doivent parvenir ensemble à une solution satisfaisante, idéalement comme s'il s'agissait d'un expert humain.

## ◆ Applications

Ce projet de thèse se fonde sur des résultats récents et originaux, concernant la modélisation de l'argumentation humaine. Il s'agit donc de transférer ces résultats théoriques vers les applications.

L'innovation qui est recherchée dans cette thèse est propre à améliorer les usages sur la Toile et sur les systèmes informatiques en général, chaque fois qu'une argumentation en langue naturelle aura une valeur ajoutée. Nous disposons d'un atout sur le plan théorique, qui provient du modèle d'argumentation que nous avons développé sur la base de la plausibilité cognitive. L'idée, en proposant cette thèse, est de profiter de cet atout pour proposer une avancée sur la conception des agents conversationnels.

Les agents conversationnels sont de plus en plus utilisés dans le commerce électronique, où ils montrent déjà certains avantages (augmentation de l'attractivité des sites, diminution du délai d'obtention des informations dû à la simplicité de formulation en langue naturelle ; augmentation du temps de séjour sur le site, lié au caractère récréatif du dialogue ; réduction du nombre d'abandons de transaction ; élévation du taux de transformation (ratio acheteurs/visiteurs)). La technique qu'il est proposé de développer dans le cadre de cette thèse est de nature à améliorer chacune de ces performances.

Nous envisageons des collaborations avec des entreprises identifiées engagées dans le développement d'agents conversationnels.

#### ◆ Contexte de réalisation

TELECOM ParisTech concentre un ensemble de compétences qui constituera un environnement scientifique et technique idéal pour le doctorant sur ce thème de recherche. La thèse sera accueillie au département Informatique et Réseaux, dans le groupe IC2 (Interaction, Cognition et Complexité). Elle s'inscrit au cœur du projet structurant MILC (Modélisations interdisciplinaires du langage et de la cognition).

#### ◆ Bibliographie

Airenti, G., Bara, B. G. & Colombetti, M. (1993). Conversation and behavior games in the pragmatics of dialogue. *Cognitive Science*, 17, 197-256.

Amgoud, L., Herzig, A. & Mercier, D. (2003). Calcul des intentions d'agent à partir de ses désirs. In B. Chaib-Draa & A. Herzig (Eds.), *Actes des journées francophones 'Modèles formels de l'interaction' (MFI-03)*. Toulouse: Cépaduès Editions.

Bentahar, J., Moulin, B. & Chaib-Draa, B. (2003). Vers une approche pour la modélisation du dialogue basée sur les engagements et les arguments. In B. Chaib-Draa & A. Herzig (Eds.), *Actes des journées francophones 'Modèles formels de l'interaction' (MFI-03)*. Toulouse: Cépaduès Editions.

Bratman, M. E., Israel, D. J. & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4 (4), 349-355.

Chiappe, D. L. & Kukla, A. (1996). Context Selection and the Frame Problem - Commentary on Sperber and Wilson's Précis of Relevance. *Behavioral and Brain Sciences*, 19 (3), 529-530.

Dessalles, J-L. (1990). The simulation of conversations. In T. Kohonen & F. Fogelman-Soulié (Eds.), *Proceedings of the Cognitiva-90 Symposium*, 483-492. Amsterdam: North Holland.  
[http://www.enst.fr/~jld/papiers/pap.conv/Dessalles\\_90072003.html](http://www.enst.fr/~jld/papiers/pap.conv/Dessalles_90072003.html)

Dessalles, J-L. (1990). Computer assisted concept learning. In D. H. Norrie & H-W. Six (Eds.), *Proceedings of the International Conference on Computer Assisted Learning (ICCAL-90) - Lecture Notes in Computer Science 438*, 175-183. Berlin: Springer-Verlag.  
[http://www.enst.fr/~jld/papiers/pap.eao/Dessalles\\_90061501.html](http://www.enst.fr/~jld/papiers/pap.eao/Dessalles_90061501.html)

Dessalles, J-L. (1998). The interplay of desire and necessity in dialogue. In J. Hulstijn & A. Nijholt (Eds.), *Proceedings of the Workshop on Formal Semantics and Pragmatics of Dialogue (TWENDIAL-98) - Twente Workshop on Language Technology (TWLT-13)*, 89-97. Enschede: University of Twente.  
[http://www.enst.fr/~jld/papiers/pap.conv/Dessalles\\_98021201.pdf](http://www.enst.fr/~jld/papiers/pap.conv/Dessalles_98021201.pdf)

Dessalles, J-L. (2001). La récursivité dans le dialogue argumentatif. In B. Chaib-Draa & P. Enjalbert (Eds.), *Actes des journées francophones 'Modèles formels de l'interaction' (MFI-01)*, 49-59. Toulouse: Université de Toulouse.  
[http://www.enst.fr/~jld/papiers/pap.conv/Dessalles\\_01053001.pdf](http://www.enst.fr/~jld/papiers/pap.conv/Dessalles_01053001.pdf)

Dessalles, J-L. (2006). A structural model of intuitive probability. In D. Fum, F. Del Missier & A. Stocco (Eds.), *Proceedings of the seventh International Conference on Cognitive Modeling*, 86-91. Trieste, IT: Edizioni Goliardiche.  
[http://www.enst.fr/~jld/papiers/pap.cogni/Dessalles\\_06020601.pdf](http://www.enst.fr/~jld/papiers/pap.cogni/Dessalles_06020601.pdf)

- Dessalles, J-L. (2008). A computational model of argumentation in everyday conversation: a problem-centred approach. In *Second international conference on Computational Models of Argumentation (COMMA-2008)*, to appear. Toulouse: IRIT.  
[http://www.enst.fr/~jld/papiers/pap.conv/Dessalles\\_07111204.pdf](http://www.enst.fr/~jld/papiers/pap.conv/Dessalles_07111204.pdf)
- Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77, 321-357.  
<http://dli.iit.ac.in/ijcai/IJCAI-93-VOL2/PDF/003.pdf>
- Hobbs, J. R. (2004). Abduction in natural language understanding. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics*, 724-741. Malden, MA: Blackwell Publishing.
- Hobbs, J. R., Stickel, M. & Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63 (1-2), 69-142.
- Hulstijn, J. (2000). Dialogue games are recipe for joint action. In *Proceedings of the 4th Workshop on Formal Semantics and Pragmatics of Dialogue (Gotalog-00)*, 99-106. Göteborg: Gothenburg Papers in Computational Linguistics 00-5.  
<http://www.ling.gu.se/konferenser/gotalog2000/FinalP/joris.ps>
- Lee, M. & Wilks, Y. (1996). *Speech acts on demand*. Mexico City: Proceedings of the 5th International Pragmatics Conference.
- Magnani, L. (2001). *Abduction, reason and science - Processes of discovery and explanation*. New York: Kluwer Academic.
- Maudet, N. (2001). *Modéliser l'aspect conventionnel des interactions langagières : la contribution des jeux de dialogue*. Thèse de doctorat en informatique, Université P. Sabatier, Toulouse.
- Muntigl, P. & Turnbull, W. (1998). Conversational structure and facework in arguing. *Journal of Pragmatics*, 29 (3), 225-256.
- Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, 32, 57-95.
- Shapiro, S., Pagnuco, M. & Lespérance, Y. (2000). Iterated belief change in the situation calculus. In *Proc. of the 7th Conference on Principles on Knowledge Representation and Reasoning (KR2000)*. Morgan Kaufman Publishers.
- Tallon, J-M., Vergnaud, J-C. & Zamir, S. (2003). Contradicting beliefs and communication: the static case. In B. Chaib-Draa & A. Herzig (Eds.), *Actes des journées francophones 'Modèles formels de l'interaction' (MFI-03)*. Toulouse: Cépaduès Editions.