Direction des Relations Internationales (DRI)

Programme INRIA "Equipes Associées"

I. DEFINITION

EQUIPE ASSOCIEE	*InToHyLo*
sélection	**2009**

Equipe-Projet INRIA : Traitement Automatique des Langues : Représentations, Inférences et Sémantique (TALARIS)	Organisme étranger partenaire : Grupo de Lógica y Computabilidad (GLyC)
Centre de recherche INRIA : INRIA Nancy, Grand Est Thème INRIA : SYM	Pays : Argentina

	Coordinateur français	Coordinateur étranger
Nom, prénom	ARECES, Carlos	FIGUEIRA, Santiago
Grade/statut	CR1 INRIA	Professor and CONICET Researcher
Organisme d'appartenance *(précisez le département et/ou le laboratoire)*	TALARIS, LORIA	GLyC, Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
Adresse postale	INRIA Nancy Grand Est TALARIS, Bat. B 615, rue du Jardin Botanique 54600 Villers lès Nancy Cedex France	Departamento de Computación, Facultad de Ciencias Exactas y Naturales. (C1428EGA) Pabellón I, Ciudad Universitaria, Buenos Aires, Argentina
URL	http://talaris.loria.fr http://www.loria.fr/~areces	http://www.glyc.dc.uba.ar http://www.glyc.dc.uba.ar/santiago
Téléphone	+33 (0)3 54 95 84 90	+54 11 4576 3390/6 int. 705
Télécopie	+33 (0)3 83 41 30 79	+54 11 4576 3359
Courriel	carlos.areces@loria.fr	sfigueir@dc.uba.ar

La proposition en bref

Titre de la thématique de collaboration *(en français et en anglais)* :

English: Inference Tools for Hybrid Logics and Applications for Natural Language Processing (InToHyLo)

Francais: Services d'Inférence pour les Logiques Hybrides et leurs Applications au le TAL

Descriptif *(environ 10 lignes)* :

The main aim of the InToHyLo project (Inference Tools for Hybrid Logics and Applications for Natural Language Processing) is to investigate inference methods for hybrid logics, to develop highly optimized inference tools based on these methods, and to use these tools in natural language applications. Talaris and GLyC are currently leaders in automated theorem proving for hybrid logics, and they are the developers of the two provers HyLoRes (based on resolution) and HTab (based on tableaux). With the InToHyLo project we want to investigate how to combine resolution and tableaux algorithms to allow our provers to collaborate and share partial results. We will integrate our tools in a platform suitable for inference in NLP applications (focusing on Dialogue Systems and Textual Entailment). This platform will include not only tools for satisfiability testing, but also for model building, model checking, bisimulation checking, and knowledge maintenance and retrieval. Finally, we want to develop parallel inference algorithms to improve performance, and distributed testing to speed up developing.

Présentation détaillée de l'Équipe Associée

1. Objectifs scientifiques de la proposition *(1 à 2 pages)*

The main aim of the InToHyLo project (Inference Tools for Hybrid Logics) is to investigate inference methods for hybrid logics, to develop highly optimized inference tools based on these methods, and to use these tools in natural language applications with a special focus on Dialogue Systems and Textual Entailment. Talaris and GLyC have been working in collaboration in this area since 2003. As a result of this collaboration they are seen nowadays as leaders in the area of computational hybrid logic and automated inference for hybrid logics, and they are the developers of two advanced automated theorem provers – HyLoRes, based on resolution, and HTab, based on tableaux – and of a variety of related tools (e.g., a graphical interface to the reasoners called HyLoRun, the model checker for hybrid logic Mcheck, the random formula generator Hgen, an automated testing environment, etc.).

In the framework of the InToHyLo (Inference Tools for Hybrid Logics, http://hylo.loria.fr/intohylo) project that we will carry out during the next three years, we plan to integrate the tools that we have already developed in a common platform, investigate their complexity and develop and implement optimized algorithms for new reasoning tasks (e.g., model generation, model checking, bisimulation checking, retrieval). We also want to investigate how to combine resolution and tableaux algorithms to allow our provers to collaborate (share partial results) while working on a given problem. Finally, we want to develop parallel inference algorithms to improve performance, and distributed testing to speed up developing.

Talaris and GLyC have been collaborating in theoretical research and in the development of inference algorithms and tools for hybrid logics, with particular interest in providing inference services for natural language processing (NLP) applications. During 2007, the tools developed by the teams have reached an advanced level of maturity, both in terms of their theoretical underpinnings and with respect to implementation (they are currently the best inference tools for hybrid logic available worldwide). The moment is ready for tackling their development into an integrated inference system that can be used in complex NLP applications. We are starting to integrate these tools in the InToHyLo platform (http://hylo.loria.fr/intohylo) to offer a varied spectrum of inference services for different hybrid logics.

Scientific Task envisioned for the next three years: Besides continuing their ongoing research on inference in hybrid logics, during the next three years the collaboration between the Talaris and GLyC aims to work in a number of concrete improvements and important research questions (full details are provided in Section II. Programme de Travail):

- Currently our tools treat "one problem at a time": if the same inference task is repeatedly

submitted to the reasoner, it will be each time solved from scratch. We will integrate the different tools we have developed into a server/client architecture. The inference algorithm hosted in the server will be able to reuse information from previous calls from the client;

- Develop optimizations that take advantage of this new architecture. In particular, different reasoning algorithms can collaborate and share results in a way that is transparent to the client;

- Design, implement and test parallel versions of the inference algorithms to improve performance;

- Investigate ways to enhance the range of inference services provided (e.g., model building, retrieval);

- Verify the algorithms to ensure correctness and test them for efficiency;

- Finally, we want to extend our results, algorithms and tools to alternative semantics (e.g., topological and neighbourhood semantics) and to more expressive logics (e.g., first order hybrid logics).

As mentioned above, we have already developed highly optimized provers for different hybrid logics (i.e., HyLoRes and HTab). We are currently adapting these provers to take advantage of the presence of multiple cores in a CPU. We have also adapted our testing environment to work on a grid (we are currently using Grid5000, http://www.grid5000.fr). We are currently well situated to carry out the integration of the different inference tools within a common platform and develop ways in which they can interact and share results. This will be the main goal during the first year of the project. Once the tools are integrated we will be able to test them in different NLP applications. Our main focus will be on Dialogue Systems. The main goal of the second year will be to extend the inference services provided by our tools to better match the needs of the NLP applications. In particular we will develop model building algorithms working under constraints imposed by the application, and model maintenance systems. During these phase we might need to adapt our tools to alternative semantics (e.g., topological or neighbourhood semantics) or to different hybrid languages (e.g., first-order hybrid logics, memory logics, etc.) in case the standard relational semantics or the expressive power are not suitable for the NLP task. These will be the topic of the last year of the InToHyLo project.

2. Présentation des partenaires *(1 page environ par partenaire)*

Presentation of the two teams applying for the Equipe Associée: Talaris is a specialist in the area of representation and inference in natural language processing, and some of their members (e.g., P. Blackburn, C. Areces and G. Hoffmann) actively work on theoretical and computational aspects of hybrid logics. GLyC on the other hand, specializes on computability and complexity especially for modal and hybrid logics and recently, with the addition of J. Castaño to the full research staff, is also involved in information management and information extraction. Clearly the two teams share many common interests, and at the same time complement each other in interesting ways. The following list mentions subjects relevant to InToHyLo and the members of the Talaris which are working on them: Semantics and Pragmatics of Dialogue Systems (L. Benotti, P. Blackburn, A. Denis, and M. Quignard), Textual Entailment (P. Bedaride and C. Gardent), Generation of Referring Expressions (C. Areces and G. Hoffmann), Inference Services for NLP (C. Areces, P. Bedaride, L. Benotti, P. Blackburn and G. Hoffmann), Hybrid and Memory Logics (C. Areces, P. Blackburn, and G. Hoffmann). GLyC members are world specialists on inference algorithms, complexity and computability for hybrid logics and, on the other hand, actively working on computational linguistics and information extraction (in particular, treating problems that require some context sensitive information, parsing, lexical information, temporal reasoning and modelling, event change, and extraction of complex relations from scientific texts.) The following list mentions subjects relevant to InToHyLo and the members of GLyC which are working on them: Expressive Power of Hybrid and Memory Logics (S. Figueira, S. Mera), Inference Methods (D. Gorín and D. Koile), Parallel and

Distributed Proving and Testing (D. Gorín, M. Pérez Rodríguez and D. Rubinstein), Information Maintenance and Information Extraction (C. Bruzzoni and J. Castaño), Computability and Complexity (V. Becher and R. Rodríguez).

Short CV of the researcher in charge of the Equipe Associée on the Talaris side: Carlos Eduardo Areces (http://www.loria.fr/~areces, born May 21st, 1971, in Argentina) is an INRIA researcher (CR1) in Talaris (http://talaris.loria.fr).

  Education: 1997-2000: PhD at the Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands. Promotor: Prof. Dr. Johan van Benthem. Supervisor: Dr. Maarten de Rijke. *1995-1996:* Degree of Licenciado en Ciencias de la Computación (Masters in Computer Science) at the University of Buenos Aires.

  Publications in the last two years: 1) Areces, C., Figueira, S. and Mera, S. Completeness Results for Memory Logics, in Proceedings of LFCS'09, USA, 2009. 2) Areces, C. and Figueira, D. Which Semantics for Neighbourhood Semantics. Submitted to IJCAI'09, USA, 2009. 3) Areces, C., Figueira, D., Figueira, S., and Mera, S. Expressive Power and Decidability for Memory Logics. In Proceedings of WoLLIC'08, 2008. 4) Areces, C. and Demri, S., editors. Methods for Modalities 2007, ENTCS, 2008. 5) Areces, C. and Goldblatt, R., editors. Advances in Modal Logic, Volume 7, College Publications, 2008. 6) Areces, C. and Gorín, D.. Resolution with Order and Selection for Hybrid Logics. Submitted to Journal of Automated Reasoning, 2008. 7) Areces, C., Koller, A., and Striegnitz, K.. Referring Expressions as Formulas of Description Logic. In Proceedings of the 5th INLG, USA, 2008. 8) Areces, C., Hybrid Logics: The Old and the New. In Proceedings of LogKCA-07, Spain, 2007. 9) Sustretov, D., Hoffmann, G., Areces, C., and Blackburn, P., Experiments in Theorem Proving for Topological Hybrid Logic. In Proceedings of Methods for Modalities 5, France, 2007. 10) Hoffmann, G. and Areces, C.. HTab: A Terminating Tableaux System for Hybrid Logic. In Proceedings of Methods for Modalities 5, France, 2007.

Other full research members of Talaris involved in the Equipe Associée: Patrick Blackburn (DR2 INRIA, http://www.loria.fr/~blackbur), Claire Gardent (DR2 CNRS, http://www.loria.fr/~gardent), Matthieu Quignard (CR2 CNRS, http://www.loria.fr/~quignard).

Students from Talaris involved in the Equipe Associée: 3 PhD Students (P. Bedaride, http://www.loria.fr/~bedaride, L. Benotti http://www.loria.fr/~benottil and G. Hoffmann http://www.loria.fr/~hoffmang) together with probably 2 or 3 master students from the Erasmus Mundus Masters in Language and Communication Technologies hosted by the Université Henri Poincare and the Université Nancy 2.

Short CV of the researcher in charge of the Equipe Associée on the GLyC side: Santiago Daniel Figueira (http://www.glyc.dc.uba.ar/santiago/, born March 15th, 1976, in Argentina) is the director of the Grupo de Lógica y Computabilidad (GLyC, http://www.glyc.dc.uba.ar/) and will be the Coordinator of the Argentinean side of the Equipe Associée. He is professor at the Computer Science Department (FCEyN, University of Buenos Aires) and Assistant Researcher of CONICET (http://www.conicet.gov.ar) Argentina.

  Education: 2002-2006: PhD at the Department of Computer Science, FCEyN, University of Buenos Aires. Supervisor: V. Becher. Co-supervisor: A. Nies (University of Auckland). 1995-2001: Degree of Licenciado en Ciencias de la Computación (Masters in Computer Science) at the Universidad de Buenos Aires.

  Publications in the last two years: 1) C. Areces, S. Figueira and S. Mera. Completeness Results for Memory Logics, in Proc. of LFCS'09, USA, 2009. 2) C. Areces, D. Figueira, S. Figueira and S. Mera. Expressive power and decidability for memory logics. In WoLLIC 2008 (Workshop of Logic, Languages, Information and Computation), volume 5110 of Lecture Notes in Computer Science, pages 56-68, 2008. 3) S. Figueira, D. Gorín and R. Grimson. On the formal semantics of IF-like logics. In WoLLIC 2008 (Workshop of Logic, Languages, Information and Computation), volume 5110 of Lecture Notes in Computer Science, pages 164-178, 2008. 4) B. R. Callejas Bedregal and S. Figueira. On the computing power of fuzzy Turing machines. Fuzzy Sets and Systems, 159(9):1072-1083, 2008. 5) S. Figueira, A. Nies and F. Stephan. Lowness properties and approximations of the jump. Annals of Pure and Applied Logic, 152(1-3):51-66, 2008. 6) V. Becher,

S. Figueira and R. Picchi. Turing's unpublished algorithm for normal numbers. Theoretical Computer Science, 377(1-3):126-138, 2007. 7) S. Figueira, F. Stephan and G. Wu. Randomness and universal machines. Journal of Complexity, 22(6):738--751, 2006. 8) B. Callejas Bedregal and S. Figueira. Classical computability and fuzzy Turing machines. In LATIN 2006: Theoretical Informatics, volume 3887 of LNCS, pages 154--165, 2006. 9) S. Figueira, A. Nies and F. Stephan. Lowness properties and approximations of the jump. In Workshop of Logic, Languages Information and Computation, volume 143 of ENCS, pages 45--57, 2006. 10) V. Becher, S. Figueira, S. Grigorieff and J. Miller. Randomness and halting probabilities.

Other full research members of GLyC involved in the Equipe Associée: V. Becher (Professor, http://www-2.dc.uba.ar/profesores/becher/), J. Castaño (Professor, http://www.cs.brandeis.edu/~jcastano/) and R. Rodríguez (Professor, http://www-2.dc.uba.ar/profesores/ricardo/).

Students from GLyC involved in the Equipe Associée: 2 PhD Students in co-tutelle (D. Gorín http://glyc.dc.uba.ar/daniel/ and S. Mera http://www.glyc.dc.uba.ar/sergio/), 1 PhD Student currently in LSV Cachan (D. Figueira http://www.lsv.ens-cachan.fr/~figueira/), together with at least 4 students from the Masters in Computer Science of the Universidad de Buenos Aires.

Short history of collaboration between the two groups: C. Areces is a graduate from the Departamento de Computación of the Universidad de Buenos Aires (UBA). GLyC was born in part from his continuous collaboration with the UBA during his postdoctoral studies. Dr. Areces, Dr. Blackburn and Dr. Becher are currently co-directing (as a co-tutelle between the Universidad de Buenos Aires and the Université Henri Poincare) the doctoral studies of D. Gorín and S. Mera, two research members of GLyC. Moreover, C. Areces has supervised the master thesis of four master students, members of GLyC.
   Previous exchanges between the two teams: S. Figueira visited Talaris during 2007 for three months with funds provided by Erasmus Mundus (he collaborated with C. Areces and P. Blackburn in the preparation of the course Introduction to Computability for the Erasmus Mundus Masters in Language and Communication Technologies). V. Becher has visited Talaris twice for short periods (to discuss general collaboration plans between the two teams) with GLyC funds. D. Gorín and S. Mera have visited Talaris in four occasions (twice using INRIA internship funding, once with GLyC funds, and once using funds from the Programme de Coopération MINCyT - INRIA/CNRS, the duration of the visits are between 1 and 3 months). M. Perez Rodríguez is visiting Talaris for three months for an internship funded by INRIA and the PARLO Project (http://talc.loria.fr/dokuwiki/doku.php?id=operations:parlo) within the TALC framework (TALC, Traitement Automatique des Langues et des Connaissances, is one of the five themes represented in the CPER MINS, Contrat Plan Etat Region Modélisation, Information et Systèmes Numériques, involving INRIA Nancy Grand Est, http://talc.loria.fr/).
C. Areces visits GLyC yearly (own funds, and GLyC and UBA funds). He has taught three Master courses ("Computational Modal Logics," "Description Logics" and "Advanced Modal Logics") during these visits and many invited seminars. P. Blackburn has visited GLyC once to teach a masters course ("Introduction to Hybrid Logics") together with C. Areces in 2008 (visit funded by the Microbio project and by Talaris). C. Gardent has visited GLyC once to teach a seminar on "Inference in NLP" (visit funded by Talaris and GLyC). L. Benotti has visited GLyC once for a week (own and GLyC funds).
   The two groups have developed in collaboration two automated theorem provers for hybrid logics (HyLoRes, based on the resolution method, and HTab, based on the tableaux method) and a number of other related tools (a random formula generator for testing, a testing environment, a model checker, a library of optimized functions, etc.). Information about these tools can be found in the InToHyLo portal (http://hylo.loria.fr/intohylo).
   Joint Publications: The follorwing is a list of joint publications between members of the two groups, organized by type.

- *Journals:* 1). Areces, C. and Gorín, D. Resolution with Order and Selection for Hybrid Logics. Submitted to the Journal of Automated Reasoning, 2008.

- *Conference Proceeding and Papers in Collections (refereed):* 1). Areces, C. and Figueira, D. Which Semantics for Neighbourhood Semantics. Submitted to the International Joint Conference in Artificial Intelligence, 2008. 2). Areces, C., Figueira, D., Figueira, S., and Mera, S.. Expressive Power and Decidability for Memory Logics. In Logic, Language, Information and Computation, pp. 56–68, Springer Berlin / Heidelberg, LNCS series, 2008. Proceedings of WoLLIC 2008. 3). Areces, C., Figueira, S., and Mera, S.. Completeness results for memory logics. In Proceedings of LFCS'09, the Symposium on Logic Foundations of Computer Science, LNCS series, Deerfield Beach, Florida, USA, 2009. 4). Areces, C. and Gorín, D.. Ordered Resolution with Selection for H(@). In Proceedings of LPAR 2004, pp. 125--141, Springer, Uruguay, 2005. 5). Areces, C. and Becher, V., Iterable AGM Functions, In Rott, H. and Williams, M., (eds.), Frontiers of Belief Revision, pp. 261--277. Kluwer Academic Publishers. 2001. 6). Areces, C. and Becher, V., Update, the Infinite Case. 7). Areces, C. and Becher, V. In Proc. of WAIT'99, Argentinian Workshop on Theoretical Computer Science, Buenos Aires, Argentina. 1999. 8). Areces, C., Becher, V. and Ferro, S., Characterization Results for d-Horn Formulas, In Logic, Language and Computation, Volume 3. Cavedon, L., Blackburn, P., Braisby, N. and Shimojina, A. (eds.), pp. 47--64, 1998. 9). Areces, C., Becher, V. and Ferro, S., Characterization Results for d-Horn Formulas, In Proceedings of ITALLC'98. Information Theoretic Approaches to Logic, Language and Computation. Hsi-Tow, Taiwan, 1998. 10). Areces, C. and Becher, V., On the Logic For Utopia, In the Annals of the 1st CACiC, Bahía Blanca, Argentina, 1995.

- *Master Thesis:* 1) Koile, D.. Master Thesis of the Universidad of Buenos Aires, "Computational Complexity of Sub-boolean Hybrid Logics." Supervised by Areces, C. and Gorín, D., 2008. 2) Figueira, D.. Master Thesis of the Universidad of Buenos Aires, "Bisimulation for Neighbourhood Semantics." Supervised by Areces, C., 2006. 3) Mendez, J.. Master Thesis of the Universidad of Buenos Aires, "Description Logics with Parametric Model Checking." Supervised by Areces, C.. 2005. 4) Gorín, D.. Master Thesis of the Universidad of Buenos Aires, "Resolution with Order and Selection for Hybrid Logics." Supervised by Areces, C., 2004.

- *Student Sessions:* 1). Areces, C. and Gorín, D., Hybrid Layering. Extended version of the paper in the Proc. of the ESSLLI Student Session 2005, Spain, 2006. 2). Areces, C. and Mera, S., Hybrid Logics with Concrete Domains. Extended version of the paper in the Proc. of the ESSLLI Student Session 2005, Spain, 2006.

Summary of links to people, research laboratories, organisms, etc.:

- Representative of the Equipe Associée in Talaris: Carlos Areces http://www.loria.fr/~areces

- Representative of the Equipe Associée in GLyC: Santiago Figueira http://www.glyc.dc.uba.ar/santiago

- Webpage of Talaris: http://talaris.loria.fr

- Webpage of GLyC: http://www.glyc.dc.uba.ar

- Webpage of InToHyLo: http://hylo.loria.fr/intohylo

- Link to page about Argentina of the DRI: http://www-direction.inria.fr/international/Page_Pays.php?pays=Argentine

3. Impact *(1 page maximum)*

Impact of the collaboration on the scientific objectives of the involved teams: During previous years, the collaboration between Talaris and GLyC was done mainly via internet. As the collaboration did not have any particular funds assigned, it was difficult to formulate a plan for the long terms

activities of the two teams. The only regular personal exchange were the yearly visits of C. Areces (usually with own funds and during vacation time taken in Argentina). All other visits were organized taking advantage of funds obtained from very different sources (INRIA Internships, Erasmus Mundus Master Program, Escuela de Ciencias Informaticas de la Universidad de Buenos Aires, Programme de Coopération MINCyT - INRIA/CNRS, CPER MISN Axe TALC, etc).

Even under these conditions, the research collaboration has advanced extraordinarily and we are currently leading the area of inference for hybrid logics. But the time is now ripe to organize a long term work agenda. Such agenda contains both important theoretical questions to tackle and serious development and implementation of the inference tools created by the two teams. Most importantly, the fundamental tools are now ready for their application to different Natural Language Processing task, but carrying out these items involves important new challenges that can only be met by pulling together very tightly the areas of expertise of both teams. Such an agenda can only be met if sufficient funds for regular collaboration, the hosting of meetings and exchange of ideas can be ensured for at least the next three years.

With the recent incorporation of J. Castaño to GLyC, the moment seems to be particularly propitious for the task, as his experience greatly reinforces the interest and skills of GLyC in NLP. As we have discussed in the previous pages, the topics addressed by the collaboration touch on the main interest of many full time researchers and PhD students in both groups. In particular, it will pull together the research contributions of five different PhD students. We believe that the impact of the collaboration will be enormous for both groups.

Impact of the collaboration on the relation between the teams and between the institutes involved:
The collaboration between GLyC and Talaris has by now a long story (to witness this it is enough to review the list of *more than fifteen* joint publications). Till now, this collaboration has been carried out without any particular funding framework that would help defining long terms plans, and a more effective research strategy. We believe that, starting from the results obtained up to now, the two teams are ready to make a real contribution to the field of automated inference in general, and to the automated inferences required in different natural processing task, in particular.

The inference tools of the InToHyLo platform are open source and can be freely used and further developed by anyone interested. Their use is not restricted to natural language applications: even though we are specially interested in developing inferences services for NLP, fundamental inference services like satisfiability testing, model checking, model building, etc. have very general applications. With respect to the developed tools then, InToHyLo will have impact in many other INRIA projects, and in particular in other INRIA projects interested in inference in NLP (e.g., Calligramme, Orpailleur, Signes, etc). Our theoretical results and inference methods are relevant to other CNRS and INRIA teams working in computational logic (PAREO, CAPP, LILaC, etc).

In addition to the points mentioned above, which are directly related to the main goals of InToHyLo, we are planning to take this opportunity to encourage the development of the area of Natural Language Processing in Argentina, which is currently seriously lacking. Finally, during the collaboration with GLyC, Talaris is interested in starting new joint work with other groups in Argentina. In particular, we are interested in developing collaboration with the Tools and Foundations for Software Engineering Lab at the Universidad de Buenos Aires; and with the Grupo de Procesamiento de Lenguaje Natural of the Universidad Nacional de Córdoba.

4. Divers : toute autre information que vous jugerez utile d'ajouter.

A) I am not aware of ongoing projects between the Departamento de Computación de la Universidad de Buenos Aires (UBA) and INRIA. This Department is, arguably, the most important Theoretical Computer Science research centre in Argentina. I know that a number of INRIA researchers (including myself) have taught courses at the ECI (Escuela de Ciencias Informáticas, http://www.dc.uba.ar/events/eci/). Also, a number of students from the Departmento de Computación de la Universidad de Buenos Aires (including 4 members of GLyC) have visited INRIA Nancy Grand Est using the INRIA Internship Program. Other senior researchers from UBA visit LORIA sporadically. Summing up, a de facto program of exchange and collaboration is already in effect, but it has not yet been provided a suitable framework for long term strategic planning of research activities.

INRIA has provided signs of interest in collaboration with Argentina and most generally with Latin America in the form of the STIC-Argentine (now called MINCyT, for collaboration between INRIA/CNRS and Argentina) and the STIC-AmSud program (for the collaboration between Francia, Argentina, Brasil, Chile, Perú and Uruguay links INRIA with the Ministerio de Ciencia, Tecnología e Innovación Productiva (http://www.mincyt.gov.ar) MinCyT, de Argentina). C. Areces takes part of the Project Microbio which was accepted by STIC-AmSud and started in January 2008. This project links Talaris with the Grupo de Procesamiento de Lenguaje Natural de la Universidad Nacional de Córdoba.

The only project "Equipes Associées" with a partner in Argentina that I have found is the project FIM (http://www.irisa.fr/vista/ftp/fim/dossier_fim_2006.html) with the Facultad de Ingeniería (Engineering Faculty) of the Universidad de Buenos Aires.

B) The major part of the budget of 2009 is devoted to exchanges between GLyC and Talaris. Notice though that many members of Talaris organize seminars (both at the Master and PhD level) during their visits to GLyC. In addition, during the visit of senior or PhD members from GLyC, research talks are organized in the framework of Talaris regular meetings, or in different ongoing scientific events. If additional funds are secured then some of the funds provided by the Equipe Associée could be used to organize a national meeting in Buenos Aires (as we mentioned above we are interested in developing contacts with the Universidad Nacional de Córdoba and the Universidad Nacional de Rosario). In any case, Talaris will be hosting the Hybrid Logic Workshop during 2009 (the main conference on hybrid logics) and the participation of GLyC researches in this event is, of course, highly relevant. During 2010 we are planning to organize only two rounds of exchanges and use part of the funds to organize a workshop on Inference for NLP hosted by Talaris.

C) We emphasize the importance of the InToHyLo Equipe Associée as it will serve as the framework for:

- Collaboration between top level senior researchers in the fields of hybrid logics, automated inference, and representation and reasoning in natural language processing.

- The doctoral studies of two GLyC members: D. Gorín and S. Mera in co tutelle between the Universidad de Buenos Aires and the Université Henri Poincare.

- The doctoral studies of three Talaris members: P. Bedaride, L. Benotti and G. Hoffmann.

- The master thesis of at least five students in GLyC and Talaris (during the three years of the collaboration).

II. PREVISIONS 2009

Programme de travail

Description du programme scientifique de travail *(1 à 2 pages maximum)*

Talaris and GLyC have been collaborating since 2003 in theoretical research and in the development of inference algorithms and tools for hybrid logics, with particular interest in providing inference services for natural language processing (NLP) applications. During 2007, the tools developed by the teams have reached an advanced level of maturity, both in terms of their theoretical underpinnings and with respect to implementation (they are currently the best inference tools for hybrid logic available worldwide, and outperform, in certain tasks, tools like FaCT++ one of the most advanced theorem provers in the area of description logics). The moment is ready for tackling their development into an integrated inference system that can be used in complex NLP applications. We are starting to integrate these tools in the InToHyLo platform (http://hylo.loria.fr/intohylo) to offer a varied spectrum of inference services for different hybrid logics. Besides continuing our ongoing

research on inference in hybrid logics, during the next three years we aim to work in a number of concrete improvements and important research questions:

- Currently our tools treat "one problem at a time." In particular, if the same inference task is repeatedly submitted to the reasoner, it will be each time solved from scratch. Up to now, our tools have no notion of "information state" and cannot reused previous results. We want to integrate the different tools we have developed into a server/client architecture, and move from a "case by case" reasoning model into a "information state based" reasoning model".

- Develop optimizations that take advantage of this new architecture. In particular, different reasoning algorithms can collaborate and share results in a way that is transparent to the client.

- Design, implement and test parallel versions of the inference algorithms to improve performance. Some of the algorithms we use could benefit from the existence of shared memory and multiple cores in a CPU. We have already a prototype version of HyLoRes that perform resolution via parallel thread that share a common pool of solved clauses.

- Investigate ways to enhance the range of inference services provided (e.g., model building, retrieval). Both HyLoRes and HTab can now do model building when the input formula is satisfiable. But this algorithms are not yet properly optimized, and they cannot guarantee any additional property over the generated model (e.g., minimal size). We are currently exploring algorithms for retrieval. We also have prototypes for both model checkers and bisimulation checkers for different hybrid logics.

- Verify the algorithms to ensure correctness and test them for efficiency. Because we have developed different tools that can solve the same inference problem in different ways, we can test their correctness via redundancy: if we obtain the same answer from the different reasoning tools, our confidence on the correctness of the algorithms increases. But testing is time consuming. We have adapted our automated testing environment to make use of the facilities of the Grid5000 networks. This work is still very preliminary, but once completed we will be able to do distributed testing obtaining an enormous speed up of the testing process.

- Finally, we want to extend our results, algorithms and tools to alternative semantics (e.g., topological and neighbourhood semantics) and to more expressive logics (e.g., first order hybrid logics).

With the *InToHyLo* project, Talaris and GLyC aim to continue their investigation and implementation of inference algorithms for hybrid logics. During the last years, the groups have developed two independent automated provers HyLoRes (based on resolution) and HTab (based on tableaux), together with a graphical interface to both of them that is called HyLoRun. HyLoRes is the most mature system of the two, and includes many state-of-the-art optimizations. Starting from the resources and the experience obtained during the development of HyLoRes, we were able to develop the first version of HTab through 2007 and we are currently including optimizations. The next step is to take advantage of the inherent dual behaviour existing between the resolution based and the tableaux based calculi: while the resolution method performs better on unsatisfiable formulas (during resolution we only need to derive the empty clause to detect that a formula is unsatisfiable), the tableau method performs better on satisfiable formulas (while constructing a tableau we only need to find a saturated open branch to detect that a formula is satisfiable). We have already tested this intuition empirically: there are satisfiable formulas for which the current version of HTab (even when only minimal optimizations are activated) performs better than HyLoRes.

Our first step will be to transform HyLoRes and HTab into server applications, while HyLoRun will act as a client application which will connect to the provers submitting queries and displaying the results. HyLoRun will detect whether HyLoRes and/or HTab are running as servers and connect to them using either HTTP or TCP services. This architecture is the one used by the description logic provers like RACER and it has some important benefits:

- To start with, the different components of InToHyLo (initially, the two provers and the front-end) can evolve independently without interfering with applications making use of them, as long as the communication protocols are maintained. In addition, new inference tools can be added as additional servers and only the front-end will need to be modified to offer these additional services.

- Secondly, we want to investigate ways in which the two provers can collaborate while working in a given problem. With this idea in mind, we want the provers to be able to exchange information (i.e., partial results) in a manner that is transparent to the user.

- But the most important reason for choosing this architecture is that it lets us implement a notion of "information state". This idea is again a fundamental characteristic of DL provers like RACER: the user should be able to load a problem into the system, and then query it for answers. Many different queries are usually posed to the prover about the same problem, and the prover can take advantage of previous results to answer future queries.

After this first step, the reasoner's architecture will be in place to tackle the other goals mentioned above. The items that we are planning to approach during 2009 are: development of parallel algorithms, distributed testing using the Grid5000 first, and then the Grid being installed by TALC, and the investigation of additional inference services (mainly model building for hybrid logics).

By the end of 2009 we are planning to start using InToHyLo in NLP applications currently being investigated by Talaris (mainly, dialogue systems and text entailment). Tuning the inference system for these applications will be the main research topic for 2010.

Summing up, a direct result of the Equipe Associée project will be a strengthening of the ongoing, and up to now very successful, research collaboration between Talaris and GLyC. Moreover, a concrete aim during the coming years is to develop the InToHyLo platform and put it to use in different NLP projects being carried out by GLyC and Talaris. Finally, as a by product, we expect to encourage the development of the area of Natural Language Processing in Argentina, which is currently seriously lacking.

Responsabilities: To clarify how the work described above will be carried out, we list below (in order of their relevance to the activities to be carried out during 2008) the names of the main participating researchers, their areas of expertise, and their main role in InToHyLo:
Talaris: C. Areces (computational logic, complexity theory, modal logic, algorithms, he is one of the main developers of HyLoRes), P. Blackburn (modal logic, complexity theory, computational semantics, he has developed the tableaux algorithms being used in HTab), G. Hoffman (computational logic, main developer of HTab), C. Gardent (computational linguistics, generation, use of inference in NLP, she is in charge of the NLP applications making use of the InToHyLo platform), L. Benotti (dialogue systems, use of inference in NLP), P. Bedaride (textual entailment, use of inference in NLP).
GLyC: S. Figueira (computability, complexity theory, algorithms, he works on expressive power and non standard inference tasks), V. Becher (computability and complexity theory, knowledge representation, theoretical results on computability theory), D. Gorín (computational logic, algorithms, he is one of the main developers of HyLoRes), S. Mera (computational logic, complexity theory, he works on expressive power and non standard inference tasks), D. Figueira (modal logic, expressive power, alternative semantics), J. Castaño (computational linguistics, information extraction), R. Rodríguez (multivalued and fuzzy logics, non standard inference tasks).

Programme d'échanges avec budget prévisionnel

1. Echanges

The following exchanges are scheduled for 2008/2009

- In December 2008 C. Areces and L. Benotti will visit GLyC (C. Areces visit will be partially funded by the STIC-Argentine project, L. Benotti visit will be partially funded by the Microbio STIC-AmSud project). A seminar on "Memory Logics" will be organized at the Departamento de Computación, UBA. Moreover, C. Areces will teach a course on "Automated Theorem Proving" at the Universidad Nacional de Córdoba. Some of the on-going Master Thesis being co-supervised by members of Talaris and GLyC will be defended during this period. The main topic of this exchange will be investigating new optimizations that take advantage of the client/server architecture, and ways in which such an architecture can be used in a dialogue system.

- In June 2009 a group of researchers from GLyC will visit Talaris. The group will be composed by the senior researchers V. Becher and S. Figueira, the two PhD. Students D. Gorín and S. Mera, and the master student D. Koile. Senior researchers will visit for 1 month each while the two PhD students and the master student will visit Talaris for 3 months each. The visit will overlap with a number of important conferences in the area of computational logic, where the new results obtained by the two teams will be presented. In particular, Talaris will be hosting the Hybrid Logic Workshop in June. In the end of July, the European Summer School in Logic, Language and Information (http://esslli2009.labri.fr) will be hosted in Bordeaux and D. Gorín, S. Mera, and D. Koile will be attending courses there which are important for their studies. The main topic of this exchange will be to investigate alternative reasoning services for hybrid logics.

- In December 2009 C. Areces and the Talaris PhD. students L. Benotti and G. Hoffmann will visit GLyC. C. Areces will organize a one week seminar on "Knowledge Representation in NLP" at the UBA. The main topic of this exchange will be the use of inference tools in dialogue systems.

1. ESTIMATION DES DÉPENSES EN MISSIONS INRIA VERS LE PARTENAIRE	Nombre de personnes	Coût estimé
Chercheurs confirmés	2	5000 euros
Post-doctorants	0	0 euros
Doctorants	3	4500 euros
Stagiaires	0	0 euros
Autre (précisez) :	0	0 euros
Total		9500 euros

2. ESTIMATION DES DÉPENSES EN INVITATIONS DES PARTENAIRES	Nombre de personnes	Coût estimé
Chercheurs confirmés	2	5000 euros
Post-doctorants	0	0 euros
Doctorants	2	6500 euros
Stagiaires	1	3500 euros
Autre (précisez) :	0	0 euros
Total		15000 euros

2. Cofinancement

Both Talaris and GLyC have independent funding of their own obtained from their hosting institutions and other projects (see the respective webpages), but funds specifically set aside for their

collaboration are very small.
C. Areces has obtained funds from STIC-AmSud for the project Microbio
(http://www.microbioamsud.net/) to be done in collaboration with the Universidad Nacional de
Cordoba in Argentina. By coordinating missions with this project, Microbio can provide funds to
cover some of the expenses of visits of Talaris members to GLyC during 2008/2009. In particular, the
visit of L. Benotti to GLyC, planned for December 2008 is funded by coordinating the trip with a
meeting of the Microbio project in Cordoba during the same period.
C. Areces has obtained funds for exchange of researchers in the framework of the Programme de
Coopération MINCyT - INRIA/CNRS (funds provided by INRIA, CNRS and MinCyT). By
coordinating missions with this project some travel expenses can be covered. In particular, the visit of
C. Areces to GLyC, planned for December 2008 is partially funded by MINCyT-INRIA/CNRS.
Talaris has submitted a topic for the ongoing call for INRIA Internships.
Research funds in Argentina are very limited. GLyC has obtained sporadic funds to cover travel
expenses. In general, we **cannot** expect symmetric funding to match those provided by INRIA.

ESTIMATION PROSPECTIVE DES CO-FINANCEMENTS	
Organisme	Montant
INRIA Internship (3 months x 900 Euros / 2)	1350 euros
MINCyT-INRIA/CNRS	2830 euros
STIC-AmSud	4000 euros
TALC Theme in CPER MISN	3000 euros
Budget EPI	1000 euros
Total	9350 euros

3. Demande budgétaire

Indiquez, dans le tableau ci-dessous, le coût global estimé de la proposition et le budget demandé à
la DRI dans le cadre de cette Equipe Associée (maximum 20 K€).

Commentaires	Montant
A. Coût global de la proposition (*total des tableaux 1 et 2 : invitations, missions, ...*)	24500 euros
B. Cofinancements utilisés (*financements autres que Equipe Associée*)	9350 euros
Financement "Équipe Associée" demandé (A.-B.) (maximum 20 K€)	15150 euros

© INRIA - mise à jour le 11/08/2008