

Sujet: Vers des solutions génériques aux problèmes d'intégration, d'extension et de rénovation de gros systèmes de TAO ou de TALN en général.

Mot-clés: Ariane-Y, environnement de développement de TAO, programmation multilingue, LSPL, génie linguiciel, génie logiciel.

Contexte scientifique: Le projet [Ariane-Y](#) en cours consiste à réaliser un environnement de programmation linguistique complet pour la traduction automatique, en portant le système [Ariane-G5](#) existant, en unifiant ses composants, et en l'étendant. [Les 5 LSPL](#) (langages spécialisés pour la programmation linguistique), dont les différents composants « linguiciels » (variables, formats, dictionnaires, grammaires...) sont maintenant traités par un seul compilateur, qui produit une forme intermédiaire en XML.

En ce qui concerne l'EDL (environnement de développement linguiciel), on cherche à développer une interface web unique permettant de développer des systèmes de TALN hétérogènes. Leurs composants pourront être écrits dans divers LSPL, dont les compilateurs et/ou les moteurs pourront n'être disponibles que sur des sites distants. [Le projet WICALE](#) [1] a déjà produit une interface générique permettant l'échange de composants linguiciels (dictionnaires, grammaires...), de données (textes, arbres...) et de commandes avec des EDL distants (Ariane-G5 et PILAF). Un développement en cours dans le M2R de Nguyen H. T. est d'y ajouter des possibilités de navigation, à la Doxygen, en construisant des représentations HTML à partir de la forme intermédiaire XML.

Dans le futur, on souhaite améliorer les LSPL d'Ariane-Y et ceux qui auront été intégrés via WICALE dans deux directions. L'une est de les simplifier pour permettre leur utilisation par des non-spécialistes de TALN [2], comme cela est possible avec TTEDIT de J.-C. Durand, dérivé de ROBRA, ou avec l'outil de syllabification de C. Del Vigna et V. Berment [4].

L'autre est de les étendre à la programmation classique, par exemple pour pouvoir écrire des compilateurs classiques avec des outils de TA.

L'objectif qui se dessine est double. D'abord, transformer WICALE en un EDL générique complet, implémenté comme un service web, et permettant le développement mutualisé de composants linguiciels par des communautés à la fois de spécialistes et de non spécialistes, un peu comme l'encyclopédie wikipedia. L'intégration d'un nouveau LSPL et de son EDL consisterait à écrire un transformateur du LSPL en XML, et à interfacer avec l'API de l'EDL. Ensuite, produire à partir de chaque LSPL accessible via WICALE une famille de LSPL de divers degrés de 'complexité conceptuelle' (ou un LSPL unique paramétrable par un niveau de complexité), et aussi rendre sensibles aux développeurs « naïfs » les concepts sous-jacents aux (variantes des) LSPL utilisés.

Points scientifiques à étudier:

- Intégration : description uniforme et multiniveau de systèmes de TAO différents ('canevas' pour décrire l'anatomie de tels systèmes), et recherche d'une méthode générique permettant d'en tirer des 'instrumentations', premier pas vers la réingénierie de tels systèmes.
- Extension des LSPL : approfondir les travaux de M. Lafourcade sur l'extension du LSPL LT à la programmation 'usuelle', pour proposer et expérimenter une méthode applicable à tous les LSPL.
- Hiérarchisation de la complexité conceptuelle d'un LSPL : il s'agit, à partir d'un LSPL à structures de contrôle complexes comme ROBRA, d'en dériver un ou plusieurs LSPL plus simples, ou de le paramétrer par un niveau de complexité (possibilités visibles ou cachées), pour faciliter l'appropriation incrémentale des concepts sous-jacents, et l'utilisation immédiate des niveaux les plus simples.

- [1]. *Virginie Carpena, Nguyen Hong-Thai, Spécification de WICALE version 2, Document interne ,GETA, CLIPS, 2005.*
- [2]. *Valérie Bellynck, Introduction d'une vue textuelle synchronisée avec la vue géométrique primaire dans Cabri-II, thèse UJF Grenoble, Octobre 1999, 250 p.*
- [3]. *Mathieu Lafourcade, Génie logiciel pour le génie linguiciel, thèse UJF Grenoble, Décembre 1994, 289 p.*
- [4]. *Vincent Berment, Méthodes pour informatiser les langues et les groupes de langues « peu dotées », thèse UJF Grenoble, Mai 2004, 277 p.*
- [5]. *Christian Boitet, Cahier des charges d'ARIANE-Y, 1999.*
- [6]. *Hong-Thai Nguyen, Vers un 'méta-EDL' puis un 'EDL générique' pour la TAO, mémoire M2R UJF Grenoble, juin 2005, 128p.*