Facultés Universitaires Notre-Dame de la Paix – Namur

Faculté d'Informatique

Application des Machines à Etats Finis en Synthèse de la Parole

Sélection d'unités non uniformes et Correction orthographique

Dissertation présentée par

Richard Beaufort

en vue de l'obtention du grade de

Docteur en Sciences, orientation Informatique

4 mars 2008

Composition du jury

Prof. Guy Deville, ELV, FUNDP, Namur	co-promoteur
Prof. Thierry Dutoit, TCTS, FPMs, Mons	co-promoteur
Prof. Christian Fluhr, INSTN, CEA, Saclay, France	examinateur
Prof. Jean-Marie Jacquet, Faculté d'Informatique, FUNDP, Namur	promoteur
Prof. Pierre-Yves Schobbens, Faculté d'Informatique, FUNDP, Namur	examinateur
Prof. Wim Vanhoof, Faculté d'Informatique, FUNDP, Namur	président du jury

Application des Machines à Etats Finis en Synthèse de la Parole

Sélection d'unités non uniformes et Correction orthographique

Richard Beaufort

Jean-Marie Jacquet (*promoteur*) Guy Deville (*co-promoteur*) Thierry Dutoit (*co-promoteur*)

Résumé

Les machines à états finis sont des outils puissants : équivalents informatiques des langages réguliers et des relations régulières, elles s'accompagnent d'algorithmes d'optimisation et peuvent être exprimées sous la forme d'expressions régulières et de règles de réécriture. Présentes depuis l'apparition de l'informatique et abondamment employées dans les domaines les plus variés, les machines à états finis ont cependant été délaissées en traitement automatique de la langue, domaine où les chercheurs ont préféré recourir à des outils plus adaptés à la description syntaxique.

La recherche a cependant récemment réalisé un véritable bond en avant dans le domaine des machines à états finis pondérées, proposant de nouveaux algorithmes qui ont ouvert la voie à une nouvelle technologie du langage, capable de modéliser ce degré d'incertitude qui est indispensable à certains domaines du traitement automatique de la langue.

La synthèse de la parole est l'un de ces domaines. Le processus de synthèse de la parole, qui consiste à produire de la parole à partir du texte, regorge de difficultés, empreintes d'un certain degré d'incertitude, qui peuvent utilement profiter du potentiel expressif des outils à états finis dans leur ensemble.

Dans le cadre de cette thèse, deux de ces difficultés ont particulièrement retenu notre attention : l'étape de sélection d'unités non uniformes qui précède la génération de la parole, et la correction orthographique réalisée dans le contexte de l'analyse morpho-syntaxique.

Les travaux réalisés s'articulent autour de trois objectifs distincts. D'une part, nous avons désiré prouver que les machines à états finis facilitent la conception de tâches complexes sous la forme d'une succession de tâches simples. D'autre part, nous avons voulu proposer des approches nouvelles et originales en sélection d'unités non uniformes et en correction orthographique. Enfin, nous avons tâché de contribuer au domaine des machines à états finis, en définissant de nouvelles méthodes d'implémentation et de nouveaux algorithmes qui ont été intégrés dans nos propres outils : une bibliothèque de machines à états finis, et un compilateur de dictionnaires, de langages et de règles de réécriture.

Table des matières