

Note de lecture

Rubrique préparée par Denis Maurel

Université François Rabelais Tours, LI (Laboratoire d'informatique)

Stéphane CHAUDIRON, Khalid CHOUKRI, L'évaluation des technologies de traitement de la langue : les campagnes Technolangue, *Hermes-Lavoisier*, 2008, 235 pages, ISBN 978-2-7462-1992-2.

Lu par **Thierry POIBEAU**

LIPN

L'ouvrage L'évaluation des technologies de traitement de la langue : les campagnes Technolangue paru chez Hermès science et édité par Stéphane Chaudiron et Khalid Choukri vise à rendre compte des campagnes Technolangue ayant eu lieu entre 2003 et 2006, pour l'évaluation des technologies du langage.

La question de l'évaluation des technologies des langues est un problème ancien dans la mesure où les premières évaluations datent du début des années 1960 (test de Cranfield, en recherche d'information). Cette problématique se pose cependant de manière particulièrement cruciale aujourd'hui car le développement de la documentation électronique (et notamment d'Internet) a considérablement renouvelé les besoins. Il existe maintenant plusieurs entreprises offrant des technologies susceptibles de répondre aux besoins exprimés dans le monde industriel (que ce soit en recherche ou en extraction d'information ou bien encore en traduction automatique, etc.). Au-delà de la documentation électronique, les technologies de la parole ont aussi développé des outils opérationnels qu'il convient de mesurer et d'évaluer (pour des tâches telles que la reconnaissance du locuteur ou la transcription de la parole...).

Sur ce plan, l'Europe est cependant en retrait par rapport aux États-Unis qui ont depuis longtemps vu l'intérêt de mesurer régulièrement les progrès des technologies de l'information, selon différents points de vue. Une véritable « culture de l'évaluation » s'est fait jour, depuis les années 1980 dans le monde anglo-saxon, à travers des campagnes sur différents sujets (extraction d'information avec les campagnes MUC – *Message Understanding Conferences*, recherche d'information avec les campagnes TREC – *Text Retrieval Conferences*, transcription de la parole avec les campagnes *Broadcast News Transcription Evaluation*). Lancée en 2003, l'initiative Technolangue visait de manière ambitieuse à évaluer les technologies de la langue, prioritairement en français, dans huit domaines différents à la fois.

Après une préface de Joseph Mariani et un chapitre sur l'évaluation, l'ouvrage traite des campagnes suivantes : ARCADE II (alignement), CESART (acquisition de ressources terminologiques), CESTA (traduction automatique), EASY (analyseurs syntaxiques), EQueR (questions-réponses), ESTER (transcription enrichie d'émissions radiophoniques), EvaSy (synthèse de la parole à partir du texte) et MEDIA (compréhension dans les systèmes de dialogue).

Les intitulés sont explicites et montrent la variété des campagnes considérées. Chaque chapitre décrit une campagne donnée, de la définition de la tâche à la publication des résultats, en passant par l'élaboration des corpus de tests, des métriques utilisées et des discussions entre participants.

De ce point de vue, le livre contient un ensemble d'informations précieuses pour quiconque souhaiterait organiser une campagne ou simplement y participer. La lecture du livre donne une bonne idée des questions qui se posent, notamment pour obtenir un consensus parmi les participants pour une tâche donnée, fournir des résultats cohérents et assurer leur diffusion. Il faut toutefois souligner que le but du livre n'est pas en soi d'entrer dans le détail de chaque campagne. On ne trouvera donc pas de description des systèmes ayant participé ou des techniques mises en œuvre ; de même, le descriptif des tâches est souvent réduit. Ceci est sans doute inévitable pour donner un aperçu de chacune des campagnes, mais le résultat obtenu est de fait un peu répétitif : le livre se donne à lire, plus comme une référence par rapport aux campagnes Technolanguage, que comme une réflexion sur l'évaluation elle-même.

Au final, que peut-on dire de la maturité des outils de traitement des langues ? On aurait pu souhaiter un bilan critique plus poussé des différentes campagnes : on devine, en filigrane, que certaines campagnes ont été plus fructueuses que d'autres, ne serait-ce qu'en observant le nombre de participants. Suivant le domaine considéré, les tâches ont été plus ou moins difficiles à définir ; idem pour les métriques. Les articles discutent parfois de ces aspects (par exemple, au chapitre 3, sur la façon d'évaluer les extracteurs de termes), mais une discussion d'ensemble aurait été la bienvenue.

Il n'en reste pas moins que ce livre, rédigé par de très bons spécialistes des questions abordées, rend compte de manière très riche d'un ensemble de campagnes qui ont mobilisé l'essentiel de la communauté française du traitement automatique des langues et des technologies de la parole. Il en ressort donc un ouvrage riche qui fera référence pour l'évaluation en France dans l'ensemble de ces domaines.