National Research Council Canada | Conseil national de recherches Canada

Centre de recherche en technologies langagières
Language Technologies Research Centre

Colloquium / Conférence

SPEAKERS/CONFÉRENCIERS

George Foster and Roland Kuhn

Senior Researchers / chercheurs senior
National Research Council of Canada / Conseil national de recherches Canada

The NRC's Portage system at the NIST "Open MT 2012" evaluation / Le système Portage du CNRC à l'évaluation « Open MT 2012 » du NIST

Thursday, June 21, 2012
11:00 a.m.
Université du Québec en Outaouais
Alexandre-Taché campus
Room F0129
Language: English with questions accepted in both French and English

ABSTRACT

The US National Institute of Standards and Technology (NIST) periodically organizes evaluations of machine translation systems. This year, the NRC's Portage system participated in two tracks of this evaluation: Chinese-to-English and Arabic-to-English. Results of the evaluation will be officially released by NIST on August 31; we are not authorized to release detailed information ahead of this date. However, it can be revealed that the Portage system, according to the preliminary scores, placed among the very top systems for both language pairs.

In this talk, we will describe the conditions for the Open MT evaluation and the big team effort that led to the final versions of Portage submitted to the Chinese-English and Arabic-English tracks. We will also describe some of the techniques that led to the excellent performance of Portage in this evaluation. In particular, we will discuss lattice MIRA tuning and Hierarchical Lexicalized Distortion Models (HLDMs), which had a particularly large impact on performance. Finally, we will discuss some of the possible future improvements of Portage.

Speakers' Biography

George Foster is a senior researcher at the National Research Council of Canada. He is on the board of AMTA, and the editorial boards of Computational Linguistics and Machine Translation. His research has mainly focused on applications for translation technology, beginning with tools for translators, and evolving as statistical MT has become more viable. His doctoral work led to the TransType project on interactive MT via sentence completion. In 2003 he led a JHU CLSP workshop on confidence estimation for SMT. More recently he has worked on adaptation and discriminative estimation.

Roland Kuhn is a senior researcher at the National Research Council of Canada. After studying mathematical biology at the University of Toronto and the University of Chicago (where he explored computer simulation as a tool for studying the evolution of DNA), Roland developed an interest in natural language. In 1993, he received his Ph.D. in Computer Science from McGill University, with a thesis on applying decision trees to the understanding of spoken phrases. In the course of his research career, Roland has studied a diverse set of problems in natural language processing, including automatic speech recognition, machine dialogue, speaker verification/identification, speech understanding, letter-to-sound systems, phoneme-based topic spotting, and most recently, machine translation. He has contributed new ideas to several of these areas, including the cache language model for speech recognition and eigenvoices for speaker adaptation and speaker verification/identification.

The NRC/LTRC colloquium does not require advance registration, and attendance is free-of-charge.
Open to the public

Le jeudi 21 juin 2012
11 h
Université du Québec en Outaouais
Campus Alexandre-Taché
Local F-0129
Langue : anglais, mais les questions seront acceptées aussi bien en français qu'en anglais

RÉSUMÉ

Le National Institute of Standards and Technology (NIST) des États-Unis organise périodiquement des évaluations de systèmes de traduction automatique. Cette année, le CNRC a inscrit son système Portage à deux volets différents de cette évaluation : traduction chinois-vers-anglais et traduction arabe-vers-anglais. Les résultats seront officiellement divulgués par NIST le 31 août et nous ne sommes pas autorisés à fournir des informations détaillées avant cette date. Toutefois, nous pouvons dire que selon les résultats préliminaires qui nous ont été communiqués, Portage s'est placé dans le peloton de tête pour les deux couples de langues mentionnés.

Dans cette conférence, nous décrirons les conditions dans lesquelles s'effectuent ces évaluations « Open MT » ainsi que l'effort d'équipe qui nous a permis de préparer Portage pour l'un et l'autre des deux volets de notre participation. Nous fournirons également un aperçu des techniques qui ont permis à Portage de briller, comme le dispositif d'accord automatique MIRA par treillis ainsi que les Modèles de distorsion hiérarchiques lexicalisés (MDHL). Enfin, nous discuterons des améliorations futures que nous comptons apporter à Portage.

Les conférenciers

George Foster est chercheur senior au CNRC. Il est membre du conseil d'administration de l'*Association for Machine Translation in the Americas* (AMTA) et membre du comité de rédaction des revues *Computational Linguistics* et *Machine Translation*. Ses recherches se sont concentrées sur les applications en technologies de la traduction, d'abord sur des outils de traduction assistée, puis sur la traduction automatique statistique qui devient de plus en plus viable. Ses travaux de doctorat ont fait émerger le projet TransType sur la traduction interactive par complétion automatique de phrases. En 2003, il a été responsable de l'un des prestigieux ateliers d'été du CLSP de l'Université Johns Hopkins sur le thème de l'estimation de confiance en traduction automatique. Ses recherches les plus récentes portent sur l'adaptation et sur l'estimation discriminative.

Roland Kuhn est chercheur senior au Conseil national de recherches Canada. Après avoir étudié la biologie mathématique à l'Université de Toronto et à l'Université de Chicago (où il a exploré la simulation par ordinateur en tant qu'outil pour étudier l'évolution de l'ADN), Roland s'est intéressé au langage naturel. En 1993, il a obtenu son doctorat en science informatique de l'Université McGill avec une thèse sur l'application des arbres de décision à la compréhension des phrases parlées. Au cours de sa carrière de chercheur, Roland a étudié un ensemble de problèmes divers dans le traitement du langage naturel, y compris la reconnaissance automatique de la parole, le dialogue machine, la vérification/l'identification du locuteur, la compréhension de la parole, les systèmes de la lettre au son, le repérage du sujet par phonème et, plus récemment, la traduction automatique. Il a apporté de nouvelles idées dans plusieurs de ces domaines, notamment dans le modèle de langage naturel en cache (*cache language model*) pour la reconnaissance de la parole et dans les voix propres (*eigenvoices*) pour l'adaptation au locuteur ainsi que pour la vérification/l'identification du locuteur.

Il n'est pas nécessaire de s'inscrire à l'avance à cette conférence du CNRC/CRTL, et l'entrée est gratuite.
Ouvert au public