

Eric Y. Kow

Nationalité : Américaine
Courriel : eric.kow@gmail.com
Situation actuelle : Développeur en free-lance

1 Projets en freelance

1. Computational Linguistics Ltd (depuis mai 2011) : implémentation du module de génération pour un livre électronique, au sein du projet AURA, mené par Stanford Research International (<http://www.sri.com/work/projects/aura>)
2. Well-Typed LLP (mars 2011 - août 2012) : valorisation du projet Parallel GHC (financé par Microsoft Research). Ce projet a pour but de développer et promouvoir l'utilisation des technologies de programmation parallèle en Haskell dans l'industrie. Le langage Haskell a des méthodes de programmation parallèle et concurrente relativement mûres, mais peu connues en dehors du milieu académique. En collaborant avec des partenaires industriels pour appliquer ces technologies dans leurs domaines respectifs, nous essayons d'identifier et de remédier aux obstacles freinant son adoption dans le monde industriel. Ma participation portait sur la diffusion des nouvelles et des savoirs dans la communauté Haskell. J'ai rédigé un tutoriel et des articles de blog mensuels, donnant les nouveaux développements dans le parallélisme Haskell et des introductions aux concepts utilisés. (<http://www.haskell.org/haskellwiki/Parallel/Digest>)

2 Expérience professionnelle postdoctorale

Research Fellow (chercheur contractuel) à la *University of Brighton* (équipe *Natural Language Technology Group*, temps plein de décembre 2007 à décembre 2010, puis mi-temps de décembre 2010 à août 2011)

1. Développement de méthodes et techniques de génération probabiliste et de méthodes d'évaluation (avec Anja Belz) : j'ai étudié l'utilisation des méthodes probabilistes comme moyen de développer plus rapidement des systèmes de génération en domaine restreint (e.g. la météo marine). Cette recherche nécessitait l'utilisation d'évaluations, par des êtres humains, de la qualité de textes produits, et le projet a donc abordé l'évaluation de l'efficacité des diverses méthodes d'obtention de ces jugements.
2. Campagnes d'évaluation (avec Anja Belz) : une part importante de mon travail a concerné la préparation des campagnes d'évaluation sur des diverses sous-tâches de la génération (TUNA 2008; GREC 2008, 2009 et 2010). Dans ces campagnes, je préparais les sites webs, les scripts d'évaluation, la chaîne d'automatisation de ces traitements d'évaluation, les expériences d'évaluation par des juges humains. Ce travail a abouti au développement d'un environnement logiciel nommé LG-EVAL, et de modules Perl et Haskell, pour simplifier la gestion des expériences d'évaluation.
3. Extraction d'information (avec Roger Evans) : dans le contexte du projet 3D-COFORM, j'ai développé un prototype de système d'extraction d'informations. Ce système donne un support automatisé pour la saisie manuelle des métadonnées à partir de descriptions textuelles de fouilles archéologiques. Le prototype repose sur les outils suivants : le système GATE, l'analyseur RASP, et une nouvelle composante de parcours de graphe sémantique (structures RMRS) écrit en Scala.

3 Formation doctorale

Doctorat en Informatique de l'Université Henri Poincaré (Nancy 1)

Directrice	Claire Gardent (CNRS/LORIA)
Titre	<i>Réalisation de surface : ambiguïté et déterminisme</i>
Mots-clés	Réalisation de surface, grammaires d'arbres adjoints, metagrammaires, génération
Jury	John Carroll, Professeur, Université de Sussex, Brighton (Rapporteur) Patrick Saint-Dizier, Directeur de Recherche CNRS, IRIT Toulouse (Rapporteur) Dominique Méry, Professeur, Université Henri Poincaré, LORIA Nancy (Examineur) Eric De La Clergerie, Chargé de Recherche, INRIA Rocquencourt (Examineur) Claire Gardent, Directrice de Recherche CNRS, LORIA Nancy (Directrice de thèse)

Soutenue le 14 novembre 2007

Résumé de la thèse

La réalisation de surface est une partie du processus global de génération de langue naturelle. Elle peut être vue comme l'inverse de l'analyse en ce sens où, étant donné une grammaire et une représentation du sens, le réalisateur de surface produit une chaîne en langue naturelle que la grammaire associe à un sens donné en entrée. Cette thèse présente trois extensions de GENI, un réalisateur de surface pour une grammaire de type FB-LTAG (Feature-Based Tree Adjoining Grammar).

La première extension augmente l'efficacité du réalisateur pour le traitement de l'ambiguïté lexicale. C'est une adaptation de l'optimisation par « étiquetage électrostatique » qui existe déjà pour l'analyse, qui consiste à associer les items lexicaux à des ensembles de polarités, et à éliminer les combinaisons dont les polarités ne sont pas neutres.

La deuxième extension concerne le nombre de sorties retournées par le réalisateur. En temps normal, l'algorithme GENI retourne *toutes* les phrases associées à une même forme logique. Alors qu'on peut considérer que ces entrées ont le même sens, elles présentent souvent de subtiles nuances. Il est important que les systèmes de génération contrôlent ces facteurs supplémentaires. Ici, nous montrons comment la spécification de l'entrée peut être augmentée d'annotations qui permettent un tel contrôle des sorties. L'extension est permise par le fait que la grammaire FB-LTAG utilisée par le générateur a été construite à partir d'une « métagrammaire », mettant explicitement en œuvre les généralisations qu'elle code.

La dernière extension donne la possibilité au réalisateur de servir d'environnement de débogage de la métagrammaire. Les erreurs dans la métagrammaire peuvent avoir des conséquences importantes pour la grammaire. Comme le réalisateur donne en sortie toutes les chaînes associées à une sémantique d'entrée, il peut être utilisé pour trouver ces erreurs et les localiser dans la métagrammaire.

4 Formations antérieures

1. Université Henri Poincaré, Nancy, 2003-2004

DEA en Informatique

Directrice : Claire Gardent

Sujet du mémoire : *Optimising a Surface Realiser*

2. Université de Pennsylvanie, Philadelphie, 1997-2001
BSE (équivalent Maîtrise) en Informatique
(année 2000/2001 à Université d'Édimbourg)
Mention : *cum laude*
3. Stanton College Preparatory High School, Jacksonville, 1993-1997
Diplôme de lycée, et Baccalauréat International

Écoles d'été

1. ESSLLI 2005 : École d'Été Européenne en Logique, Linguistique, et Informatique, Edimbourg, Royaume-Uni (1 semaine)
Cours suivis :
Natural language semantic representations as types (T. Fernando).
Term Rewriting Systems (F. Baader)
Natural Language Semantics (R. Muskens)
2. ESSLLI 2003, Vienne, Autriche (2 semaines)
Cours suivis :
Knowledge-Intensive Question Answering (M. de Rijke et B. Webber)
Discourse Representation Theory – an Updated Survey (A. Bende-Farkas et J. van Genabith)
DRT and Related Theories (D. Hardt)
Epistemic logic and multiagent systems (LC Verbrugge, R. et H. van Ditmarsch)
Transducers (S. Shieber)
3. ESSLLI 2002, Trento, Italie (2 semaines)
Cours suivis :
Formal Issues in Natural Language Generation (K. van Deemter et M. Stone)
Foundations of Constraint Satisfaction (R. Barták)
Shallow Text Processing (C. Brew)
Presupposition (R. van der Sandt and B. Geurts)
Working with Corpora (M. Wolters)
Modal and Epistemic Logic in Computer Science (A. R Lomuscio et W. van der Hoek)

5 Autre expérience professionnelle

1. Laboratoire LORIA, Nancy 2001-2003
Ingénieur expert (équipe Langue et Dialogue).
Conception et implémentation d'une bibliothèque d'interopérabilité des composantes dans un système de dialogue

6 Collaborations scientifiques

1. Projet européen OZONE, 2003 : Le projet OZONE avait pour but de créer un cadre général pour le développement des logiciels d'intelligence ambiante. Dans ce contexte, ma contribution a consisté à l'extension et l'intégration de la bibliothèque développée pour le projet MIAMM.

2. Projet européen MIAMM, 2001-2003 : Le projet MIAMM (Multidimensional Information Access Using Multiple Modalities), coordonné par Laurent Romary, visait la création d'un cadre général pour la conception de systèmes de dialogue multimodaux. En tant que ingénieur expert, j'ai conçu et implémenté un bibliothèque qui permet ou composants dans un système de dialogue de se communiquer à travers des standards ouverts (SOAP).

7 Animation de la communauté scientifique

1. ENLG 2011 (European Workshop on Natural Language Generation)
membre du comité de relecture
2. COLING 2010 (International conference on Natural Language Generation)
membre du comité de relecture de la Section « Summarization and Generation »
3. ACL 2008 (Association for Computational Linguistics)
membre du comité de relecture de la Session Étudiante
4. INLG 2006 (International conference on Natural Language Generation)
membre du comité de relecture de la Session Étudiante
5. LACL 2005 (Logical Aspects of Computational Linguistics)
membre du comité de relecture de la Session Étudiante
6. ESSLLI 2004 (École d'Été Européenne en Logique, Linguistique, et Informatique) ; membre du comité local d'organisation à Nancy

8 Initiatives dans le cadre du support informatique au LORIA

1. Mise en place d'un serveur TRAC (2006)
Le serveur est utilisé par le laboratoire dans la gestion de projets de développement de logiciels. Il est intégré avec le système de contrôle de versions DARCS (<http://trac.loria.fr>)
2. Mise en place d'un serveur wiki (2005)
Le wiki est utilisé par le laboratoire pour rédiger des documents collaboratifs (e.g., documentation de logiciels, foire aux questions, compte-rendus des réunions)
(<http://wiki.loria.fr>)
3. Mise en place d'un environnement d'annotation de corpus au sein du projet Langue et Dialogue (utilisé en particulier par Hélène Manuélian lors de sa thèse).

9 Enseignements dispensés

1. Introduction à XML (9h stage)
Formation continue, Nancy (15 étudiants)
2. Programmation Caml (14h CM, 20h TD, 20h TP)
DEUG 1, Épinal (40 étudiants)

10 Langues

1. anglais (langue maternelle)
2. français (parlé couramment, lu et écrit)

11 Langages informatiques

1. Haskell (utilisé activement)
2. Java, Python, Perl, C/C++, Javascript (utilisés antérieurement)

12 Publications

12.1 Conférences internationales avec comité de lecture et actes

1. *Natural Language Generation for a Smart Biology Textbook*. Eva Banik, Eric Kow, Vinay Chaudhri, Nikhil Dinesh, et Umangi Oza International Natural Language Generation Conference (INLG). Starved Rock, Illinois, États-Unis, 2012.
2. *LG-Eval: A Toolkit for Creating Online Language Evaluation Experiments*. Eric Kow et Anja Belz. Eighth International Conference on Language Resources and Evaluation (LREC). Istanbul, Turquie, 2012.
3. *Discrete vs. Continuous Rating Scales for Language Evaluation in NLP*. Anja Belz et Eric Kow. Association for Computational Linguistics (ACL). Portland, Oregon, États-Unis, 2011.
4. *Comparing Rating Scales and Preference Judgements in Language Evaluation*. Anja Belz et Eric Kow. International Natural Language Generation Conference. Dublin, Irlande, 2010.
5. *Extracting Parallel Fragments from Comparable Corpora for Data-to-text Generation*. Anja Belz et Eric Kow. International Natural Language Generation Conference. Dublin, Irlande, 2010.
6. *A Symbolic Approach to Near-Deterministic Surface Realisation using Tree Adjoining Grammar*. Claire Gardent et Eric Kow. Association for Computational Linguistics (ACL). Prague, République Tchèque, 2007.
7. *Towards Reusable NLP Components*. Amalia Todirascu, Laurent Romary et Eric Kow. Third International Conference on Language Resources and Evaluation (LREC), Las Palmas, Espagne, 2002.

12.2 Ateliers internationaux avec comité de lecture et actes

8. *Unsupervised Alignment of Comparable Data and Text Resources*. Anja Belz et Eric Kow. 4th Workshop on Building and Using Comparable Corpora (BUCC). Portland, Oregon, États-Unis, 2011.
9. *System Building Cost vs. Output Quality in Data-To-Text Generation*. Anja Belz et Eric Kow European Natural Language Generation (ENLG) Workshop. Athènes, Grèce, 2009. (*Best paper award*)
10. *Spotting Overgeneration Suspects*. Claire Gardent et Eric Kow. European Natural Language Generation (ENLG) Workshop, Dagstuhl, Allemagne, 2007.
11. *GenI: Natural language generation in Haskell*. Eric Kow. Haskell'06, Portland, États-Unis, 2006.
12. *SemTAG, the LORIA toolbox for TAG-based Parsing and Generation*. Eric Kow, Yannick Parmentier, Claire Gardent. 8th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+8). Sydney, Australie, 2006.
13. *Three reasons to adopt TAG-based surface realisation*. Claire Gardent et Eric Kow. 8th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+8). Sydney, Australie, 2006.
14. *Generating and selecting grammatical paraphrases*. Claire Gardent et Eric Kow. 10th European Workshop on Natural Language Generation (ENLG), Aberdeen, UK, 2005
15. *Which bridges for bridging definite descriptions?* Claire Gardent, Helene Manuelian et Eric Kow. 4th International Workshop on Linguistically Interpreted Corpora (LINC), Budapest, Hongrie, 2003

12.3 Campagnes d'évaluation

16. *The GREC Challenges 2010: Overview and Evaluation Results*. Anja Belz et Eric Kow. International Natural Language Generation Conference (INLG). Dublin, Irlande.
17. *The GREC Main Subject Reference Generation Challenge 2009: Overview and Evaluation Results*. Anja Belz, Eric Kow, Jette Viethen et Albert Gatt. European Workshop on Natural Language Generation (ENLG). Athènes, Grèce.
18. *The TUNA-REG Challenge 2009: Overview and Evaluation Results*. Albert Gatt, Anja Belz et Eric Kow. ENLG.
19. *The GREC Named Entity Generation Challenge 2009: Overview and Evaluation Results*. Anja Belz, Eric Kow et Jette Viethen. ENLG.
20. *The GREC Challenge 2008: Overview and Evaluation Results*. Anja Belz, Eric Kow, Jette Viethen et Albert Gatt. International Natural Language Generation Conference (INLG). Salt Fork, Ohio, États-Unis.
21. *The TUNA Challenge 2008: Overview and Evaluation Results*, Albert Gatt, Anja Belz et Eric Kow International Natural Language Generation Conference (INLG). Salt Fork, Ohio, États-Unis.

12.4 Conférence nationale avec comité de lecture et actes

22. *Une réalisateur de surface basée sur une grammaire réversible* Claire Gardent et Eric Kow, Traitement Automatique des Langues Naturelles (TALN), juin 2007

12.5 Session étudiante de conférences avec comité de lecture et actes

23. *Adapting polarised disambiguation to surface realisation*. Eric Kow. 17th European Summer School in Logic, Language and Information (ESSLI) Student session, Edinburgh, UK, 2005.