

L'apprenant de Schemas
pour la Structuration Rhétorique du
Génération Automatique du Textes

Pablo A. Duboue

Computer Science Department
Columbia University
in the city of New York



Mon doctorat

- Il me manque un an et demi pour finir mes études
 - J'ai commencé mon doctorat à Columbia en 1999.

- Mes publications

(Tous sont "Pablo A. Duboue and Kathleen R. McKeown")

EMNLP 2003 *Statistical Acquisition of Content Selection Rules for Natural Language Generation.*

INLG 2002 *Content Planner Construction Via Evolutionary Algorithms and a Corpus-based Fitness Function.*

ACL-EACL 2001 *Empirically Estimating Order Constraints for Content Planning in Generation.*

Présentation EMNLP

- En Anglais

Ma these

- Choisir le schema que maximise le suivant probabilité conditionnelle
 - $P(\text{Schema} | (\text{Donnés}, \text{Textes}))$

- Estimer ça probabilité avec trois distances:

- Le distance entre les données choisies pour le système et les données qui sont donc les textes.

$$\|K_{ordinateur} - K_{humain}\|$$

- Le distance entre les arbres rhétoriques générés et les arbres rhétoriques du textes.

$$\|R_{ordinateur} - R_{humain}\|$$

- Le distance entre les textes générés et les textes humaines.

$$\|D_{ordinateur} - D_{humain}\|$$