

# UDel: Refining a Method of Named Entity Generation

Charles F. Greenbacker, Nicole L. Sparks, Kathleen F. McCoy, and Che-Yu Kuo

Department of Computer and Information Sciences

University of Delaware

Newark, Delaware, USA

[charlieg | sparks | mccoy | kuo]@cis.udel.edu

## Abstract

This report describes the methods and results of a system developed for the GREC Named Entity Challenge 2010. We detail the refinements made to our 2009 submission and present the output of the self-evaluation on the development data set.

## 1 Introduction

The GREC Named Entity Challenge 2010 (NEG) is an NLG shared task whereby submitted systems must select a referring expression from a list of options for each mention of each person in a text. The corpus is a collection of 2,000 introductory sections from Wikipedia articles about individual people in which all mentions of person entities have been annotated. An in-depth description of the task, along with the evaluation results from the previous year, is provided by Belz et al. (2009).

Our 2009 submission (Greenbacker and McCoy, 2009a) was an extension of the system we developed for the GREC Main Subject Reference Generation Challenge (MSR) (Greenbacker and McCoy, 2009b). Although our system performed reasonably-well in predicting REG08-Type in the NEG task, our string accuracy scores were disappointingly-low, especially when compared to the other competing systems and our own performance in the MSR task. As suggested by the evaluators (Belz et al., 2009), this was due in large part to our reliance on the list of REs being in a particular order, which had changed for the NEG task.

## 2 Method

The first improvement we made to our existing methods related to the manner by which we selected the specific RE to employ. In 2009, we trained a series of decision trees to predict REG08-Type based on our psycholinguistically-inspired

feature set (described in (Greenbacker and McCoy, 2009c)), and then simply chose the first option in the list of REs matching the predicted type. For 2010, we incorporated the case of each RE into our target attribute so that the decision tree classifier would predict both the type and case for the given reference. Then, we applied a series of rules governing the length of initial and subsequent REs involving a person’s name (following Nenkova and McKeown (2003)), as well as ‘back-offs’ if the predicted type or case were not available.

Another improvement we made involved our method of determining whether the use of a pronoun would introduce ambiguity in a given context. Previously, we searched for references to other people entities since the most recent mention of the entity at hand, and if any were found, we assumed these would cause the use of a pronoun to be ambiguous. However, this failed to account for the fact that personal pronouns in English are gender-specific (ie. the mention of a male individual would not make the use of “she” ambiguous). So, we refined this by determining the gender of each named entity (by seeing which personal pronouns were associated with it in the list of REs), and only noting ambiguity when the current entity and candidate interfering antecedent were of the same gender.

Other small changes from 2009 include an expanded abbreviation set in the sentence segmenter, separate decision trees for the main subject and other entities, and fixing how we handled embedded REF elements with unspecified mention IDs.

## 3 Results

Scores for REG08-Type precision & recall, string accuracy, and string-edit distance are presented in Figure 1. These were computed on the entire development set, as well as the three subsets, using the `geval.pl` self-evaluation tool provided in the

NEG participants' pack.

While we were able to achieve an improvement of nearly 50% over our 2009 scores in string accuracy, we saw less than a 1% gain in overall REG08-Type performance.

Metric	Score
Type Precision/Recall	0.757995735607676
String Accuracy	0.650496141124587
Mean Edit Distance	0.875413450937156
Normalized Distance	0.319266300067796

(a) Scores on the entire development set.

Metric	Score
Type Precision/Recall	0.735294117647059
String Accuracy	0.623287671232877
Mean Edit Distance	0.839041095890411
Normalized Distance	0.345490867579909

(b) Scores on the 'Chefs' subset.

Metric	Score
Type Precision/Recall	0.790769230769231
String Accuracy	0.683544303797468
Mean Edit Distance	0.882911392405063
Normalized Distance	0.279837251356239

(c) Scores on the 'Composers' subset.

Metric	Score
Type Precision/Recall	0.745928338762215
String Accuracy	0.642140468227425
Mean Edit Distance	0.903010033444816
Normalized Distance	0.335326519731057

(d) Scores on the 'Inventors' subset.

Figure 1: Scores on the development set obtained via the `geval.pl` self-evaluation tool. REG08-Type precision and recall were equal in all four sets.

## 4 Conclusions

The fact that our string accuracy scores improved over our 2009 submission far more than REG08-Type prediction is hardly surprising. Our efforts during this iteration of the NEG task were primarily focused on enhancing our methods of choosing the best RE once the reference type was selected.

We remain several points below the best-performing team from 2009 (ICSI-Berkeley), possibly due to the inclusion of additional items in their feature set, or the use of Conditional Random Fields as their learning technique (Favre and Bohnet, 2009).

## 5 Future Work

Moving forward, we hope to expand our feature set by including the morphology of words immediately surrounding the reference, as well as a more extensive reference history, as suggested by (Favre and Bohnet, 2009). We suspect that these features may play a significant role in determining the type of referenced used, the prediction of which acts as a 'bottleneck' in generating exact REs.

We would also like to compare the efficacy of several different machine learning techniques as applied to our feature set and the NEG task.

## References

- Anja Belz, Eric Kow, and Jette Viethen. 2009. The GREC named entity generation challenge 2009: Overview and evaluation results. In *Proceedings of the 2009 Workshop on Language Generation and Summarisation (UCNLG+Sum 2009)*, pages 88–98, Suntec, Singapore, August. Association for Computational Linguistics.
- Benoit Favre and Bernd Bohnet. 2009. ICSI-CRF: The generation of references to the main subject and named entities using conditional random fields. In *Proceedings of the 2009 Workshop on Language Generation and Summarisation (UCNLG+Sum 2009)*, pages 99–100, Suntec, Singapore, August. Association for Computational Linguistics.
- Charles Greenbacker and Kathleen McCoy. 2009a. UDel: Extending reference generation to multiple entities. In *Proceedings of the 2009 Workshop on Language Generation and Summarisation (UCNLG+Sum 2009)*, pages 105–106, Suntec, Singapore, August. Association for Computational Linguistics.
- Charles Greenbacker and Kathleen McCoy. 2009b. UDel: Generating referring expressions guided by psycholinguistic findings. In *Proceedings of the 2009 Workshop on Language Generation and Summarisation (UCNLG+Sum 2009)*, pages 101–102, Suntec, Singapore, August. Association for Computational Linguistics.
- Charles F. Greenbacker and Kathleen F. McCoy. 2009c. Feature selection for reference generation as informed by psycholinguistic research. In *Proceedings of the CogSci 2009 Workshop on Production of Referring Expressions (PRE-Cogsci 2009)*, Amsterdam, July.
- Ani Nenkova and Kathleen McKeown. 2003. Improving the coherence of multi-document summaries: a corpus study for modeling the syntactic realization of entities. Technical Report CU-CS-001-03, Columbia University, Computer Science Department.