We use named entity (NE) classes to improve broad coverage surface realization with the OpenCCG realizer.

We collapse certain multi-word NEs and interpolate a language model where NEs are replaced by their class labels yielding the largest quality increase.

Substantial further benefit is obtained by including class information in the hypertagging (supertagging for realization) component of the system.

Our efforts yield a state-of-the-art BLEU score of 2.9173 on Section 23 of the CCGbank.

A targeted manual evaluation confirms that the BLEU score increase corresponds to a significant rise in fluency.

**Logical Forms and Hypertagging**

**Exploiting NE Classes**

Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 09.

Mr. Otero, who apparently has an unpublished number, also could not be reached.

**Class Features in Hypertagging**

We also experimented with a hypertagging model trained over the collapsed corpus, where the semantic classes of the elementary lexical predications, along with the class features of their adjacent nodes, are added as features. Class information improves performance on the collapsed corpus.

**Class-Related N-gram Models**

We created language model training data with semantic classes replacing actual words, in order to address data sparsity issues caused by rare words in the same semantic class (e.g., Pierre Vinken ⇒ PERSON).

During realization, word forms are generated, but are then replaced by their semantic classes and scored using the semantic class replaces n-gram model, similar to Oh and Rudnicky (2002).

As the specific word may matter, the class-replaced model is interpolated at the word level with an ordinary word-based language model, as well as with a factorized language model over POS tags and supertags.

**Collapsing Named Entities**

Of 2,331 NEs annotated by the BBN corpus (Weischedel and Bernstein, 2005), 238 are not realized correctly by our baseline realizer; for example, Texas Instruments Japan Ltd. is realized as Japan Texas Instruments Ltd.

Inspired by Hogan et al.'s Experiment 1, we decided to use the BBN NE corpus to collapse certain classes of NEs.

Unlike their experiment, where all the NEs annotated by the BBN corpus are collapsed, we chose to collapse into single tokens only NEs whose exact form can be reasonably expected to be specified in the input to the realizer (e.g., Pierre Vinken).

While some quantificational or comparative phrases like more than $10,000 are annotated as MONEY in the BBN corpus, in our view only $10,000 should be collapsed into an atomic unit, with more than handled compositionally according to the semantics assigned to it by the grammar.

**Discussion**

Ref-neg: 2000:10 Mr. Otero, who apparently has an unpublished number, also could not be reached.

Baseline: Mr. Otero, who apparently has an unpublished number, also could not be reached.

Best: Mr. Otero, who apparently has an unpublished number, also could not be reached.

**Targeted Manual Evaluation**

Along the lines of Callison-Burch et al. (2006), two native speakers (two of the authors) provided ratings for a random sample of 49 realizations that differed between the baseline and best conditions on the collapsed corpus. Sentences were rated in the context of the preceding sentence; ratings for a random sample of 49 realizations that differed between the baseline and best conditions on the collapsed corpus. Sentences were rated in the context of the preceding sentence. The judges were not aware of the condition (best/baseline) while doing the rating. Ratings of the two judges were averaged for each item. The increase in fluency from 3.36 to 3.61 represented a significant difference (paired t-test, 1-tailed, $p = 0.015$), while adequacy scores did not differ significantly.

**Acknowledgements**

This work was supported in part by NSF IIS-0641297 and by an allocation of computing time from the Ohio Supercomputer Center. Our thanks also to Josef VanGenabith, the OSU Chippers group and the anonymous reviewers for helpful comments and discussion.