# Automatic extraction of property norm-like features from large text corpora with gold standard, human and semantic-similarity evaluations.

*Colin Kelly[1], Barry Devereux[2] and Anna Korhonen[1]*
*[1]Computer Laboratory, University of Cambridge*
*[2]Centre for Speech, Language and the Brain, University of Cambridge*
*contact: colin.kelly@cl.cam.ac.uk*

Property norms (e.g., banana *is* **yellow,** aeroplane *has* **wings)** play a key role in cognitive science, forming the basis for many recent theoretical accounts of conceptual representations (e.g., Cree et al., 2006; Grondin et al., 2009; Randall et al., 2004). Such norms are typically derived from norming studies where a large number of human participants elicit properties for a set of concepts (e.g., McRae et al., 2005).

We propose a novel system for the automated extraction of property norm-like triples from large text corpora which uses syntactic, encyclopedic, semantic and statistical information to guide extraction. Prior work has mostly focused on the simpler task of relation extraction (e.g., Davidov et al., 2007; Pantel and Pennacchiotti, 2008), and only Devereux et al. (2009) and Baroni et al. (2009) have attempted the more ambitious task of full property extraction.

The first stage of our system employs part of speech and grammatical relation (GR) rules to extract candidate property triples in a concept *relation* **feature** format from our C&C-parsed corpora (Wikipedia and the British National Corpus). These rules are derived from typical patterns of instantiation of property norm-like information, manually generated from a subset of Wikipedia. The system makes two passes over the corpus. The first pass extracts candidate features from short GR paths. The second pass uses these candidate features and longer GR paths to generate concept-relation-feature triples. These triples are grouped and lemmatized to remove inflectional variation, giving a set of triples with their corpus production frequency.

The second stage of our system scores each triple on four statistical metrics: log-likelihood, pointwise mutual information, entropy and a semantic reweighting factor. A linear combination of scores from these metrics and normalised frequency information yields an ordered list of triples – we select the top twenty highest-scoring triples as our system output.

We evaluate our system by directly comparing our output with a synonym-expanded subset of the McRae norms (Baroni et al., 2008). Our best F-scores are 0.147 (matching on both the relation and feature portions of our triples) and 0.285 (matching on features alone) when using a combination of triples output from the BNC and Wikipedia. We also calculate F-scores using Baroni et al.'s evaluation criteria (i.e., evaluating the top ten features), outperforming their best F-score of 0.239 with an F-score of 0.321.

In a second evaluation, two human judges evaluated our top twenty output (excluding triples marked as correct by our gold standard) over 15 concepts. For our combined BNC/Wikipedia system, our judges marked 50.2% of this subset of the output triples as correct/plausible (and the remainder as wrong/wrong-but-related) indicating that the majority of our output (i.e., the top twenty, including correct, gold standard triples) tends to be correct/plausible.

Finally, we compare concept-concept similarity, calculating cosine similarity from our triples, with a WordNet semantic-similarity measure (Leacock and Chodorow, 1998) shown to correlate highly with human judgements (Budanitsky and Hirst, 2006). Our best system achieves a Pearson correlation of 0.522 with the WordNet measure when considering both features and relations, exceeding the correlation with the McRae norms themselves (0.470).

Our system offers an original and effective method for property norm-like triple extraction: our gold standard comparison shows improvement on the current state-of-the-art and subsequent evaluations demonstrate the humanlike nature of our output.

References:


Baroni, M., Evert, S., and Lenci, A., editors (2008). *ESSLLI 2008 Workshop on Distributional Lexical Semantics*.

Baroni, M., Murphy, B., Eduard, B., and Massimo, P. (2009). Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science,* pages 1–33.

Budanitsky, A. and Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics,* 32(1):13–47.

Cree, G., McNorgan, C., and McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology Learning Memory and Cognition,* 32(4):643.

Davidov, D., Rappoport, A., and Koppel, M. (2007). Fully unsupervised discovery of concept-specific relationships by web mining. In *Annual Meeting-Association For Computational Linguistics,* volume 45, page 232.

Devereux, B., Pilkington, N., Poibeau, T., and Korhonen, A. (2009). Towards Unrestricted, Large-Scale Acquisition of Feature-Based Conceptual Representations from Corpus Data. *Research on Language & Computation,* pages 1–34.

Grondin, R., Lupker, S., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language,* 60(1):1–19.

Leacock, C. and Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. *WordNet: An electronic lexical database,* 49(2):265–283.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods, Instruments, and Computers,* 37:547–559.

Pantel, P. and Pennacchiotti, M. (2008). Automatically harvesting and ontologizing semantic relations. In *Proceeding of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge,* pages 171–195. IOS Press.

Randall, B., Moss, H., Rodd, J., Greer, M., and Tyler, L. (2004). Distinctiveness and correlation in conceptual structure: Behavioral and computational studies. *Journal of Experimental Psychology Learning Memory and Cognition,* 30(2):393–406.