Automatically Summarizing Medical Literature Using Comparative Sentences Structure

Ana Lucic & Catherine Blake, Ph.D

Center for Informatics Research in Science and Scholarship (CIRSS)

Background

- Head-to-head comparisons of viable clinical alternatives are critical for Comparative Effectiveness Research (Sox, Helfand et al. 2010).
- Even in studies that don't include viable clinical alternatives, authors compare drug results to a control population.
- From a linguistic perspective comparisons have been called "almost notorious for their syntactic complexity" (Bresnan, 1973).

Goal

Our goal is to summarize literature based on the specific entities (agent and object, depicted in bold) being compared and the basis of the comparison (depicted in italics)

> In the present study, *uterine weights* of intact animals treated with **TAM** was decreased as compared with **controls**, although not significantly. 12189200

Approach

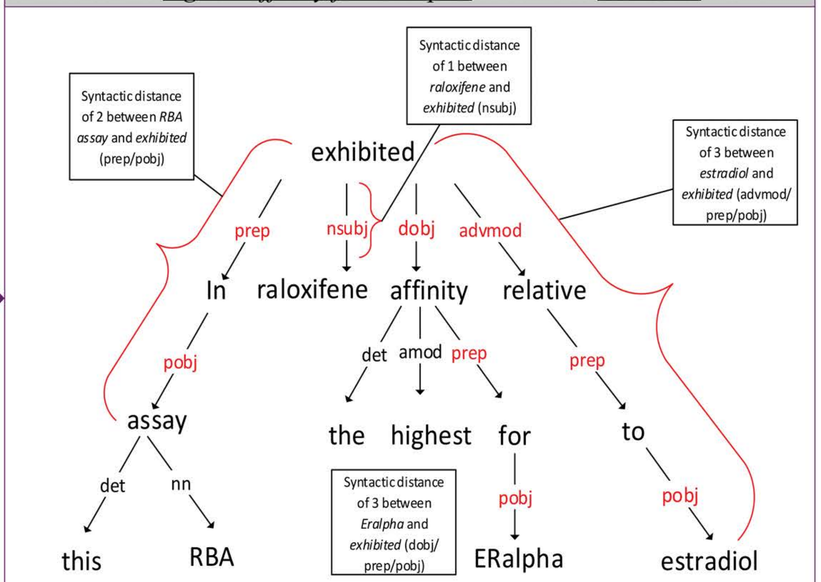
- Build on earlier work that identifies comparison sentences (Hoon Park & Blake, 2012).
- Focus on identifying specific noun phrases in comparison sentences that correspond to each of the two compared entities (agent and object) and the basis of comparison (aspect of change).
- Create features from the dependency graph representation (Stanford dependency parser version 3.2.0, de Marneffe et al., 2006).
- Apply two classifiers: support vector machine (SVM) and generalized linear model (GLM).

This poster is made possible in part by a grant from the U.S. Institute of Museum and Library Services (IMLS), Laura Bush 21st Century Librarian Program Grant Number RE-05-12-0054-12 Socio-technical Data Analytics (SODA).

Results

- Transition phrase verbs that indicate change and evidence were effective features.
- The SVM classifier worked better than GLM (or Naïve Bayes or decision trees, data not shown here).
- The second entity (the object) was easiest to identify. The first entity (agent) and the basis of change were more difficult.
- These results provide raw materials necessary to create a summary of medical literature using comparative sentences.

Figure 1: Features from the sentence "In this RBA assay, **raloxifene** exhibited the *highest affinity for ERalpha* relative to **estradiol**" 10579349



References

Bresnan, W. J. (1973). "Syntax of the Comparative Clause Construction in English." *Linguistic Inquiry* 4(3): 275-343.

de Marneffe, M., MacCartney, B., and Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. http://nlp.stanford.edu/manning/papers/LREC_2.pdf (accessed March 5, 2014).

Fiszman, Marcelo, Demner-Fushman, Dina, Lang, Francois M., Goetz, Philip, & Rindflesch, Thomas C. (2007). *Interpreting comparative constructions in biomedical text*. Proceedings of the Workshop on *BioNLP*: Biological, Translational, and Clinical Language Processing, 2007, Prague, Czech Republic.

Hoon P., D, & Blake, C. (2012). "Identifying comparative sentences in full-text scientific articles." Association of Computational Linguistics, Workshop on Detecting Structure in Scholarly Discourse, July 12, Jeju South Korea.

Sox, H. C., M. Helfand, J. Grimshaw, K. Dickersin, P. L. M. Editors, D. Tovey, J. A. Knottnerus and P. Tugwell (2010). "Comparative effectiveness research: challenges for medical journals." PLoS Med 7(4): e1000269.