

## **Filtered collocations as features in verbal polysemy**

### **disambiguation: A case study of the Chinese verb *kao* ‘bake’**

Yu-Yun Chang and Shu-Kai Hsieh

National Taiwan University

In Generative Lexicon Theory (GLT) (Pustejovsky 1995), co-composition is one of the generative devices proposed to explain the cases of verbal polysemous behavior where more than one function application is allowed. The English baking verbs were used as examples to illustrate how their arguments co-specify the verb with *qualia unification*. Some studies (Blutner 2002; Carston 2002; Falkum 2007) stated that the information of pragmatics and world knowledge need to be considered as well. Therefore, this study would like to examine whether GLT could be practiced in a real-world Natural Language Processing (NLP) application using collocations. We have conducted a fine-grained logical polysemy disambiguation task, taking open-sourced the Leiden Weibo Corpus as resource and computing with Support Vector Machine (SVM) classifier. Within the classifier, we have taken collocated verbs under GLT as main features. In addition, measure words and syntactic patterns are extracted as additional features for comparison. Our study investigates the logical polysemy of the Chinese verb *kao* ‘bake’. We find that GLT could help in identifying logically polysemous cases; additional features would help the classifier achieve a higher performance.

**Keywords:** Generative Lexicon Theory, co-composition, baking verb, logical polysemy, collocation