

A Model of Novel Compound Production

Dermot Lynott (dermot.lynott@ucd.ie)

&

Mark T. Keane (mark.keane@ucd.ie)

Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland

PUNC: Producing and Understanding Novel Combinations

The PUNC model (Producing and Understanding Novel Combinations) is the first model to capture both production and comprehension aspects of conceptual combination within the same theoretical framework. The comprehension side of PUNC has been detailed elsewhere (see Lynott, Tagalakis & Keane, 2004), so here we give a brief overview of the production side of the model.

Lynott (2004) has proposed the Integrated Production and Comprehension (IPAC) Theory of conceptual combination. The IPAC theory seeks to describe the two sides of the conceptual combination coin, comprehension and production, within the same theoretical framework. Lynott draws together several factors that have been shown to influence both sides of conceptual combination in similar ways. Central to this view are the factors of Diagnosticity, Informativeness and Plausibility (inspired by earlier work by Costello & Keane, 2000). PUNC is a computational implementation of this theory, with the model taking as input short descriptions of novel entities (e.g., “a beetle that eats cacti” or “a prickly beetle”) and, using the aforementioned factors, outputs candidate labels together with an overall acceptability score for each label. Below, we provide a brief description of the stages the model undergoes, from taking in an entity description to outputting candidate labels and assigning acceptability scores.

A description is input to PUNC (e.g., a beetle that eats cacti; a beetle that is prickly). Concepts are activated either by being explicitly mentioned in the description or through the description containing a feature that is diagnostic of another concept. So, the “is prickly” feature would also activate the concept *cactus*. Each of these concepts forms part of a set of candidate modifiers for the head concept (e.g., *beetle*). The individual concepts’ features are then activated, prioritised by their diagnosticity. For example, for the concept *cactus* “is prickly” is more diagnostic than “can conserve water” and so has greater activation. These features are used to determine whether a modifier is informative with respect to the head concept as PUNC considers each candidate modifier in turn and whether it can form part of a valid, acceptable label for the entity being described. For example, *cactus beetle* is output as a valid label for the described entity since “is prickly” is a highly diagnostic feature of the concept *cactus* and this feature is also informative with respect to the head concept *beetle*

(i.e., beetles are not by default prickly). Labels are considered informative if they incorporate some new information relative to the head concept. In this way, the informativeness of a label is a binary affair. A label such as “wood tree” meaning “a tree made from wood” would not be considered informative and so would be rejected as a possible label. As such, informativeness is a primary pragmatic constraint within the theory and model.

PUNC assigns overall acceptability scores to each of the candidate labels, based on the relative diagnosticity of the features used, the informativeness of the label and the plausibility of the relation that links the two concepts in the compound. For example, using a highly diagnostic feature of *cactus* to form a label contributes positively to the acceptability of a compound; if a less diagnostic feature had activated *cactus* the resultant score would be reduced.

Finally, the plausibility of the relation linking the head and modifier concepts contributes to the acceptability of the label. For the description “a beetle that eats cacti”, the label *cactus beetle* would be considered highly plausible since there is a reciprocal “eats” relation between the concepts – beetles can eat things, and cacti, as vegetative matter can be eaten. On the other hand, *brick beetle* as a label for “a beetle that eats bricks” is considered less plausible as bricks are not usually considered edible.

Lynott (2004) has found that by using such pragmatic constraints in an integrated fashion PUNC not only reflects people’s choice of label for novel entities, but its overall acceptability scores correlate highly with people’s ratings of how good specific compounds are as labels for entity descriptions.

Acknowledgments

The work presented here has been funded through a grant from the Irish Research Council for Science, Engineering and Technology under the Embark Initiative.

References

- Costello, F. J. & Keane, M. T. (2000). Efficient creativity: Constraints on conceptual combination. *Cognitive Science*, 24, 299-349.
- Lynott, D., Tagalakis, G. & Keane, M. T. (2004). Conceptual combination with PUNC. To appear in *Artificial Intelligence Review*.
- Lynott, D. (2004). *Comprehension and Production in Conceptual Combination*. Doctoral dissertation, University College Dublin, Ireland.