

# Word Play Generation

## Vivaek Shivakumar

### Computer Systems Lab 2009-2010

## Abstract

Computational humor is a subfield of artificial intelligence focusing on computer recognition and generation of humorous language. This paper investigates methods for generating various types of word play (e.g., puns, palindromes, acronyms) using a lexicon from the Natural Language Toolkit and the semantic web WordNet as well as phonetic information, which is the basis for many puns. Although no formal model or theory for humor exists, pun-generation has been implemented using simple, constrained models and this project attempts to recreate such implementations with possible expansions to more types of word play.

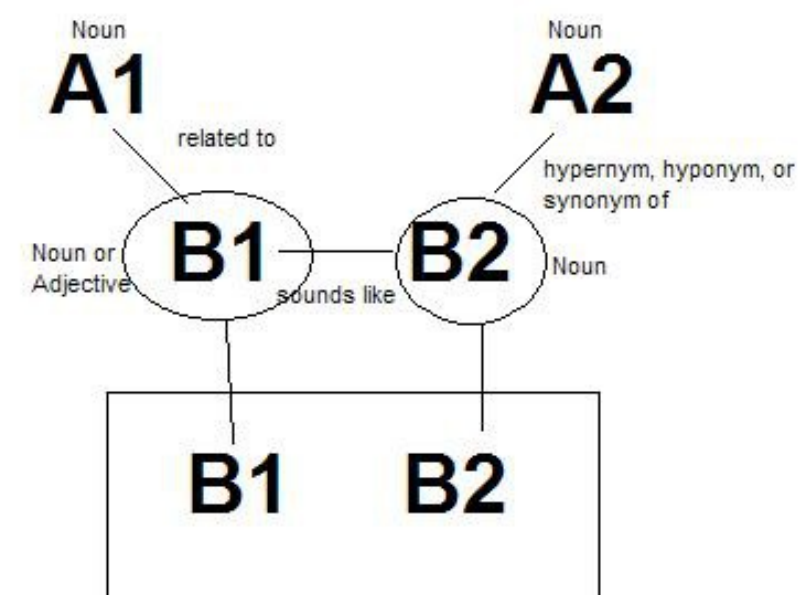


Fig 1: example schema for a "What do you get when you cross A1 and A2?" pun riddle

## Background and Introduction

Humor is a very relevant topic to artificial intelligence in the fields of natural language generation and computer creativity. The purpose of this project is to investigate and implement methods for generating various types of word play including puns, palindromes, acronyms, etc.

Humor has been studied extensively in a social or otherwise non-AI context, but no formal models or theories exist to be applied to AI. Nevertheless, the possibility of computer generation or recognition of humor is by no means impossible, as several computer programs have been created to generate small puns. These include JAPE, which was based on certain classes of puns modeled by schema and templates and utilizing the large lexicon WordNet, and HAHAcronym, which re-analyzed existing acronyms with humorous interpretations.

Examples of word play:

- What do you get when you cross a murderer with a breakfast food? A cereal killer.
- Pasteurize: too fast to see. (An example of redefinition wordplay)
- Palindrome: A man, a plan, a canal- Panama!
- Anagram: "Eleven plus two" = "Twelve plus one"

Other types of puns and jokes include knock-knock jokes and Tom Swifty puns.

## Methodology

**Punning Riddles:** Using a schema (see Fig. 1) appropriate words are picked that could potentially form a pun, based on the specified word relations. Then based on various properties of the constituent words (e.g. article requirement, animacy, compound words) a template is applied to yield the surface form of the joke.

**Palindromes:** To make palindromes consisting of valid English words, a stack and recursive algorithm is used simultaneously with an incremental segmentation process to accept or reject a given string state. Once a state is accepted, the entire string along with its segmented reverse forms the palindrome.

**Acronyms:** An input serves two purposes: to form the letters of the acronym itself, as well as providing the semantic seeds for the words that will be used to fill those letters. Both WordNet semantic relations as well as internet data on commonly used words with the input are used to provide a greater success rate as well as less nonsensical results.

## Results and Conclusions

**Punning Riddles:**

Some nonsensical or unusable results:  
"rabbit-> coney-> phoney->dissimulator"

Some workable results:

"rabbit->hare->fair->honest"

e.g., "What do you call an honest rabbit? A fair hare."

Currently, some problems to resolve include:

- NLTK's inclusion of proper nouns, British variants, and obscure words
- The surface syntactic features to be adjusted, such as when to use articles
- Classifying nouns so that only certain types of nouns (e.g. living things, ideas, places) are used in certain contexts

Ideally, the project will be able to generate (with some success rate) various types of puns that are recognizably humorous, although not necessarily funny to the user. If success is found with English, the model can be applied to other language with semantic webs.

**Palindromes:**

- race car
- red art trader
- no ill im million

**Acronyms:**

ORDER = Orderliness Rules Decree Edict Rescript  
BAD = Below Average Decency  
CIA = Collecting Independent Activities  
CIA = Collecting Intelligence Abroad  
LAW = Legal Activity \_\_\_\_\_

The apparent problem for both acronyms and palindromes is the use of extremely obscure words as well as the over-use of very common words. Also a problem is the fact that words are not picked in any order to fit a syntactic structure, which leads to nonsense.