

Project Proposal: Pun Generator

**Anonymized**

Idaho State University

### **Proposal**

Our idea for a creative computational system is a pun generator that will initially rely on two single-word inputs given by a user. The number of inputs allowed may be adjusted throughout the development process. The inputs will be altered into homophones or homographs through their phonetic spellings or conceptual meanings; new sentences will then be formed, linking together the new words and concepts. We plan to tackle the domain of puns in a way that goes beyond simple word association or rhyming outputs. The idea is to use a large dataset containing regular (non-humorous) sentences that are commonly used in the English language. In this way we can devise a system that uses context of frequently used language to connect the homophones and homographs with sentences.

### **Creativity**

In order for a system to be creative, it must be natural and based on human behavior. People manipulate language to form a play on words that uses the multiple meanings of words and similar-sounding words to produce humor. This ability is considered creative because it possesses typicality by using homophones and homographs but also includes novelty to create the story to intertwine the words together. Humans would have difficulty thinking of multiple puns within a small timeframe, which is where a computerized pun generator would be useful. In the pun generator, typicality would be achieved when finding homophones and homographs from the two word input by the user. Novelty would be accomplished by composing a sentence involving all of these words. Each pun would be novel depending on the two random word inputs from the user. The artefacts generated would be assessed for its quality and novelty using

a threshold. If an artefact is not above the threshold, the artefact is discarded. Therefore, we believe a computerized pun generator would meet the criteria for being a creative system.

### **Approach**

The domain of our computational creativity system is humor. Given the humor domain, the representation for artefacts in the domain are the individual words that compose the pun, such as homophones and homographs, as the genotypic representation and the pun itself, which is the phenotypic representation. Puns use word pairs that sound similar, called homophones, but have different meanings. Puns can also use a word that has different meanings (homographs). Homophones and homographs are the main elements that will create the funniness of the pun. The knowledge base will be scraped from ConceptNet, Word2Vec, and WordNet.

Aesthetic considerations for the artefacts in the humor domain are funniness, cleverness, surprise, timeliness, and accessibility. Will the pun make people laugh? Does the pun have wit and ingenuity? Is the pun different from what is expected? Does the pun reference current events or pop culture? Will a person get the pun?

The generator function will be using neural networks, such as RNNs, LSTM, or the Constrained Markov Model. Using lists of homophones, homographs, and ConceptNet on the two words provided by the user, a pair of homophones and a homograph will be chosen. Neural networks will be used to train the model to compose a sentence of common words and punctuation that have constrained positions of homophones and homographs.

The genotypic evaluator will rate the artefact how well the individual words within the pun agree with the pun creation rules. The phenotypic evaluator will rate the artefact on how funny the pun is by assessing the pun on quality and novelty.

## Resources

### Academic papers

[https://www.researchgate.net/publication/220356042\\_The\\_Construction\\_of\\_a\\_Pun\\_Generator\\_for\\_Language\\_Skills\\_Development](https://www.researchgate.net/publication/220356042_The_Construction_of_a_Pun_Generator_for_Language_Skills_Development)

[http://vnpeng.net/papers/pungen\\_paper.pdf](http://vnpeng.net/papers/pungen_paper.pdf)

<https://www.aclweb.org/anthology/P18-1153.pdf>

<https://pdfs.semanticscholar.org/a140/66ad6fe314773224a4ba550a09073eac4d49.pdf>

<http://computationalcreativity.net/ijwcc07/papers/ritchie-et-al.pdf>

<https://link.springer.com/content/pdf/10.1023/A:1011610210506.pdf>

### Data Sets

Homographs: [https://7esl.com/homographs/#Homographs\\_List\\_A](https://7esl.com/homographs/#Homographs_List_A)

Homophones: <http://www.singularis.ltd.uk/bifroest/misc/homophones-list.html> ,

<https://www.teachingtreasures.com.au/teaching-tools/Basic-worksheets/worksheets-english/uppe>  
[r/homophones-list.htm](https://www.teachingtreasures.com.au/teaching-tools/Basic-worksheets/worksheets-english/uppe/r/homophones-list.htm) , The-Big-List-of-Homophones.pdf

Non-humorous text: <http://norvig.com/big.txt>