

# Lexical Blends Overview

# Lexical Bends Overview

Neologisms—newly coined words or new senses of an existing word—are constantly being introduced into a language, often for the purpose of naming a new concept.

# Lexical Bends Overview

Neologisms—newly coined words or new senses of an existing word—are constantly being introduced into a language, often for the purpose of naming a new concept.

Natural language processing (NLP) system operating on recently produced text will encounter new words.

# Lexical Bends Overview

Neologisms—newly coined words or new senses of an existing word—are constantly being introduced into a language, often for the purpose of naming a new concept.

Natural language processing (NLP) system operating on recently produced text will encounter new words.

Lexical resources are often a key component of an NLP system, performance of the entire system will likely suffer due to missing lexical information for neologisms.

# Lexical Bends Overview

Neologisms—newly coined words or new senses of an existing word—are constantly being introduced into a language, often for the purpose of naming a new concept.

Natural language processing (NLP) system operating on recently produced text will encounter new words.

Lexical resources are often a key component of an NLP system, performance of the entire system will likely suffer due to missing lexical information for neologisms.

Ideally, an NLP system could identify neologisms as such, and then infer various aspects of their syntactic or semantic properties necessary for the computational task at hand.

# Lexical Bends Overview

One type of neologisms are lexical blends.

# Lexical Bends Overview

One type of neologisms are lexical blends.

Most blends are formed by combining a prefix of one source word with a suffix of another source word (brunch - breakfast and lunch).

There may be overlap in the contribution of the source words (fantabulous - fantastic and fabulous).

It is also possible that one or both source words are entirely present (jetiquette - jet etiquette).

# Lexical Bends Overview

My project is to build a system, which could correctly guess the source words from which blend are made.



# Lexical Bends Overview

My project is to build a system, which could correctly guess the source words from which blend are made.

I will use machine learning techniques to train the system.

# Lexical Bends Overview

My project is to build a system, which could correctly guess the source words from which blend are made.

I will use machine learning techniques to train the system.

However, there is one problem.

# Lexical Bends Overview

My project is to build a system, which could correctly guess the source words from which blend are made.

I will use machine learning techniques to train the system.

However, there is one problem.

A standard machine learning algorithm does not directly apply because of the structure of the problem space. In classification, we typically look for a hyperplane that separates the positive and negative training examples.

# Lexical Bends Overview

My project is to build a system, which could correctly guess the source words from which blend are made.

I will use machine learning techniques to train the system.

However, there is one problem.

A standard machine learning algorithm does not directly apply because of the structure of the problem space. In classification, we typically look for a hyperplane that separates the positive and negative training examples.

We are more interested in the ranking of the word pairs and we only need the top ranked pair of words (if blend is made from two words).

# Lexical Bends Overview

Paul Cook and Suzanne Stevenson “Automatically Identifying the Source Words of Lexical Blends in English”.

# Lexical Blends Overview

Paul Cook and Suzanne Stevenson “Automatically Identifying the Source Words of Lexical Blends in English”.

Goals:

- i) recreate Cook's system,
- ii) add more features of the blends,

# Lexical Bends Overview

A candidate set for architourist, a blend of architecture and tourist.

Archimandrite tourist

Archipelago tourist

Architect behaviourist

Architect tourist

Architectural behaviourist

Architectural tourist

Architecturally behaviourist

Architecturally tourist

Architecture behaviourist

Architecture tourist

Archives tourist

Archivist tourist

# Lexical Bends Overview

Frequency of the words:

$\text{freq}(w1) / \text{freq}(\text{prefix})$

$\text{freq}(w2) / \text{freq}(\text{suffix})$

How often one words occur after the other:

$2 \times \text{freq}(w1\ w2) / (\text{freq}(w1) + \text{freq}(w2))$

$2 \times \text{freq}(w2\ w1) / (\text{freq}(w1) + \text{freq}(w2))$

How often words occur in 5 word window:

$2 \times \text{freq}(w1, w2 \text{ in a 5 word window}) / (\text{freq}(w1) + \text{freq}(w2))$



# Lexical Blends Overview

Paul Cook and Suzanne Stevenson “Automatically Identifying the Source Words of Lexical Blends in English”.

Goals:

- i) recreate Cook's system,
- ii) add more features of the blends,
- iii) try out different machine learning techniques,
- iv) experiments with the systems using the original Cook's system as the baseline and evaluate the performance of my added features and machine learning techniques.

# Lexical Bends Overview

Overall, my goal is to improve the system, which is described in Cook's and Stevenson's paper, and have a valid system for finding source words for the blends.

My time lines for the project is:

- i) to have Cook's system by the end of this semester,
- ii) try out different features of blends and different machines learning techniques by the half of second semester,
- iii) experiment and evaluate the systems and write the report by the end of second semester.

Thank you for your time