Nehovah: A Neologism Creator Nomen Ipsum

Michael R. Smith, Ryan Hintze and Dan Ventura
Brigham Young University
Department of Computer Science
Neologisms

- Language is constantly evolving:
  - Recently added words to the Merriam-Webster Dictionary:
    - Freegan: “an activist who scavenges for free food (as in waste receptacles at stores and restaurants) as a means of reducing consumption of resources”
    - Turducken: “a boneless chicken stuffed into a boneless duck stuffed into a boneless turkey”
    - Gamification: “the process of adding games or gamelike elements to something (as a task) so as to encourage participation”
  - Especially important to many businesses
    - 433,651 trademark applications filed in 2013
    - 4.5% increase from 2012
  - Literary devices
    - Convey meaning – “cybernetics”
    - Add humor and interest as done by Dr. Seuss
Outline

- Overview of Nehovah
- Synonym and hyponym discovery
- Blending words
- Scoring
  - Word structure scoring
  - Concept scoring
  - Uniqueness scoring
  - Pop culture scoring
- Combining Scores
Overview of Nehovah

- **Goal**: Create novel words expressing the concepts from multiple source words

- **Synonym and hyponym discovery**

- **Blending words**
  - **Evaluation**
Synonym and Hyponym Discovery

- **Soda**
  - **WordNet**
    - 2 senses of soda
      - Sense 1
        sodium carbonate, washing soda, sal soda, soda ash, soda
        => salt
      - Sense 2
        pop, soda, soda pop, soda water, tonic
        => soft drink

- **Addiction**
  - **WordNet**
    - 3 senses of addiction
      - Sense 1
        addiction, dependence, dependance, dependency, habituation
        => physical condition, physiological state, physiological condition
      - Sense 2
        addiction
        => craving
      - Sense 3
        addiction
        => award, awarding
Synonym and Hyponym Discovery

- **Soda**
  - **TheTopTens**
    - Top Ten Sodas
      - Coca Cola Cherry
      - Coca Cola
      - Berries and Cream Dr. Pepper
      - ...
    - Best Soda Flavors
      - Dr. Pepper
      - Root Beer
      - Mountain Dew
      - ...
  - Top 10 Sodas Mexico vs U.S.A.
    - A&W Root Beer (U.S.)
    - Peñafiel (Mexico)

- **Addiction**
  - **TheTopTens**
    - Greatest Janes Addiction Songs
      - Ted, Just Admit It...
      - Been Caught Stealing
      - Stop
      - Mountain Song
      - ...

Validated against Wikipedia
## Blending Words

<table>
<thead>
<tr>
<th>carbonate</th>
<th>physiological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefix</td>
<td>Suffix</td>
</tr>
<tr>
<td>--</td>
<td>carbonate</td>
</tr>
<tr>
<td>ca</td>
<td>carbonate</td>
</tr>
<tr>
<td>car</td>
<td>bonate</td>
</tr>
<tr>
<td>carbo</td>
<td>nate</td>
</tr>
<tr>
<td>carbona</td>
<td>te</td>
</tr>
<tr>
<td>carbonate</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

caological      

carbological    

carbonacal     

physite         

physiolobonate  

physiologicalte
Scoring

- **Word Structure**
  - Is the neologism realistic?

- **Concept**
  - How apparent are the concepts from the source words?

- **Uniqueness**
  - How unique are the words used to create the neologism?

- **Pop Culture**
  - Are pop culture references used to create the neologism?
Word Structure Scoring

- **Valid letter sequence**
  - Create an N-gram tree
    - Any Sequence of N letters within a word
    - Useful for confirming that a word is realistic
    - Rejects neologisms with invalid letter sequences

- **Syllable exchange/Overlapping letters**
  - Does the neologism maintain the flow of one of the source words?
    - Ginormous (from giant and enormous)
    - Linsanity (from Lin and insanity)
N-Gram Tree
N-Gram Tree

- Added “Ape”
N-Gram Tree

- Added “Apply”

If letter sequence is not in the N-gram tree, deemed as invalid
Word Structure Scoring

- **Valid letter sequence**
  - Create an N-gram tree
    - Any Sequence of N letters within a word
    - Useful for confirming that a word is realistic
    - Rejects neologisms with invalid letter sequences

- **Syllable exchange/Overlapping letters**
  - Does the neologism maintain the flow of one of the source words?
    - Ginormous (from giant and enormous)
    - Linsanity (from Lin and insanity)
Word Structure Scoring

- Valid letter sequence
  - Create an N-gram tree
    - Any Sequence of N letters within a word
    - Useful for confirming that a word is realistic
    - Rejects neologisms with invalid letter sequences

- Syllable exchange/Overlapping letters
  - Does the neologism maintain the flow of one of the source words?
    - Ginormous (from giant and enormous)
    - Linsanity (from Lin and insanity)
Concept Scoring

- **MoreWords**
  Normalized FPM for words starting/end with a prefix/suffix
  - **Starting with “carbo”**
    - carbon (50)
    - carbonate (11)
    - carbohydrate (8)
    - carboniferous (4)
    - carbohydrates (2)
    - carbonates (1)
  - **Ending with “ological”**
    - psychological (28)
    - biological (20)
    - ideological (17)
    - technological (16)
    - geological (10)
    ...

- **Pop Culture**
  - Normalized count of the number of times a word appears in TheTopTens lists
Uniqueness Scoring

For each based work, the relative frequency per million words as provided by MoreWords

- pop – 79
- carbonate – 11
- tonic – 11
- soda – 8
Pop Culture Scoring

- Piece-wise equation:
  - 1 if both base words are pop culture words
  - 0.5 if only one is a pop culture word
  - 0 otherwise
Combining Scores

- A weighted average of the scores
  - Weights are assigned by a user
## Examples

<table>
<thead>
<tr>
<th>Neologism</th>
<th>Base words</th>
<th>Source Words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neohovah</strong></td>
<td>neologism</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>divinage</strong></td>
<td>divine</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>machinative</strong></td>
<td>machine</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>Spritependency</strong></td>
<td>Sprite</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>Pepsidiction</strong></td>
<td>Pepsi</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>pisome</strong></td>
<td>pie</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>pimazing</strong></td>
<td>pie</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>iniquitivate</strong></td>
<td>iniquity</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>immoralize</strong></td>
<td>immorality</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>coalesception</strong></td>
<td>coalesce</td>
<td>neologism</td>
</tr>
<tr>
<td><strong>portmanception</strong></td>
<td>portmanteau</td>
<td>neologism</td>
</tr>
<tr>
<td></td>
<td>Jehovah</td>
<td>creator</td>
</tr>
<tr>
<td></td>
<td>coinage</td>
<td>creator</td>
</tr>
<tr>
<td></td>
<td>creative</td>
<td>creative</td>
</tr>
<tr>
<td></td>
<td>dependency</td>
<td>dependency</td>
</tr>
<tr>
<td></td>
<td>soda</td>
<td>soda</td>
</tr>
<tr>
<td></td>
<td>addiction</td>
<td>addiction</td>
</tr>
<tr>
<td></td>
<td>awesome</td>
<td>awesome</td>
</tr>
<tr>
<td></td>
<td>pizza</td>
<td>pizza</td>
</tr>
<tr>
<td></td>
<td>evil</td>
<td>evil</td>
</tr>
<tr>
<td></td>
<td>school</td>
<td>school</td>
</tr>
<tr>
<td></td>
<td>concept</td>
<td>concept</td>
</tr>
<tr>
<td></td>
<td>blend</td>
<td>blend</td>
</tr>
</tbody>
</table>
Future Work

- More diverse source for synonyms/hyponyms
  - Twitter
- Improved word splitting
- Have Nehovah learn which neologisms are best
- Generate a definition for each neologism