Diachronic Analysis of Thick And Think Adjective Antonym Pair Usage In Wine Tasting Notes

Brendan FLANAGAN\textsuperscript{1} Kiyota HASHIMOTO\textsuperscript{2} and Sachio HIROKAWA\textsuperscript{3}

\textsuperscript{1} Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan
\textsuperscript{2} Faculty of Technology and Environment, Prince of Songkla University, Thailand
\textsuperscript{3} Research Institute for Information Technology, Kyushu University, Japan
E-mail: \textsuperscript{1} b.flanagan.885@s.kyushu-u.ac.jp

Abstract Antonymic adjectives in discourse can be used to express opposing properties of a meaning dimension. An example of this can be found in wine tasting notes, where the characteristics of the wine are described using a number of different dimensions. In this paper, we examine the change in use over time of antonym pairs in wine tasting notes. In particular, we analyze the change in the use of thick and thin by word co-occurrence frequency, generating visualizations for diachronic analysis.

Keywords Antonym, Winespeak, Diachronic, Mind Map, Visualization.

1. Introduction

Adjectives play the role of expressing attributes and are used frequently in wine tasting notes, often concatenated in sentences to describe complex characteristics. This use of language to describe wines in tasting notes is referred to as Winespeak. Antonym pairs that are on opposite poles of meaning dimensions are used to describe wine characteristics. In this paper, we will investigate the change in antonym use in wine tasting notes over time. The diachronic analysis presented in this paper, will target the antonym pair consisting of thick and thin, that was also the focus of a case study investigating the use within the BNC corpus in Paradis et al. \cite{1}. It was reported that use patterns relating to the width meaning dimension suggested a positive negative antonym relation, expressing opposing meanings. However, in the case of wine tasting notes, extreme degrees of the width dimension express negative qualities of wines. Therefore, it can be thought of as a double negative relation, with the middle of the dimension representing the optimal characteristic of the attribute. Firstly, we will investigate the changes in use of thick and thin by the frequency of co-occurring adjectives and the number of documents relating to the antonym pair by year. Secondly, we will examine mind map visualizations of thick and thin to analyze changes in the dissimilarity of closely related adjectives over time.

2. Data Collection

The target data for analysis is a corpus that consists of 91,010 wine tasting notes, or 255,966 sentences, that were collected from the Wine Enthusiast website\textsuperscript{1}. The attributes of each wine, such as: date when the note was published, winery, region, and grape variety were collected along with the text of the wine tasting notes. The tasting notes in the corpus were published from 1999 to 2014. An overview of the non-uniform distribution of publication dates by year can be seen in Figure 1.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Fig1.jpg}
\caption{Distribution of wine tasting notes published from 1999 to 2014.}
\end{figure}

The corpus was then indexed to construct a special use search engine using GETA\textsuperscript{2}. During indexing, words were stemmed and a list of 2488 adjectives that frequently occur in wine tasting notes was used as a mask for word selection. The raw frequencies of the top 10 frequent adjectives by years are shown in Table 1.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|}
\hline
\hline
\hline
\end{tabular}
\caption{Raw frequencies of the top 10 frequent adjectives by years.}
\end{table}

\textsuperscript{1} http://buyingguide.winemag.com/
\textsuperscript{2} http://geta.ex.nii.ac.jp/
Table 1. Occurrence frequency of top 10 adjectives.

<table>
<thead>
<tr>
<th>Year</th>
<th>Flavor</th>
<th>Fruit</th>
<th>Finish</th>
<th>Acid</th>
<th>Cherry</th>
<th>Tannin</th>
<th>Sweet</th>
<th>Dry</th>
<th>Palate</th>
<th>Ripe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>397</td>
<td>559</td>
<td>356</td>
<td>121</td>
<td>160</td>
<td>143</td>
<td>120</td>
<td>124</td>
<td>144</td>
<td>131</td>
</tr>
<tr>
<td>2000</td>
<td>718</td>
<td>850</td>
<td>674</td>
<td>249</td>
<td>383</td>
<td>227</td>
<td>235</td>
<td>221</td>
<td>313</td>
<td>197</td>
</tr>
<tr>
<td>2001</td>
<td>1400</td>
<td>1218</td>
<td>1123</td>
<td>496</td>
<td>574</td>
<td>467</td>
<td>383</td>
<td>493</td>
<td>655</td>
<td>315</td>
</tr>
<tr>
<td>2002</td>
<td>2101</td>
<td>1803</td>
<td>1272</td>
<td>729</td>
<td>712</td>
<td>890</td>
<td>628</td>
<td>730</td>
<td>731</td>
<td>525</td>
</tr>
<tr>
<td>2003</td>
<td>2441</td>
<td>1570</td>
<td>1650</td>
<td>858</td>
<td>954</td>
<td>1042</td>
<td>607</td>
<td>602</td>
<td>903</td>
<td>585</td>
</tr>
<tr>
<td>2004</td>
<td>2745</td>
<td>1950</td>
<td>1689</td>
<td>990</td>
<td>1289</td>
<td>1132</td>
<td>979</td>
<td>808</td>
<td>822</td>
<td>733</td>
</tr>
<tr>
<td>2005</td>
<td>3296</td>
<td>2182</td>
<td>1938</td>
<td>1158</td>
<td>1607</td>
<td>1192</td>
<td>971</td>
<td>1233</td>
<td>963</td>
<td>873</td>
</tr>
<tr>
<td>2006</td>
<td>3143</td>
<td>2226</td>
<td>1594</td>
<td>1293</td>
<td>1339</td>
<td>1036</td>
<td>834</td>
<td>1120</td>
<td>805</td>
<td>989</td>
</tr>
<tr>
<td>2007</td>
<td>3128</td>
<td>2370</td>
<td>1433</td>
<td>1111</td>
<td>1388</td>
<td>1015</td>
<td>917</td>
<td>888</td>
<td>702</td>
<td>885</td>
</tr>
<tr>
<td>2008</td>
<td>2973</td>
<td>2244</td>
<td>1110</td>
<td>1084</td>
<td>1221</td>
<td>1088</td>
<td>814</td>
<td>789</td>
<td>590</td>
<td>858</td>
</tr>
<tr>
<td>2009</td>
<td>1733</td>
<td>1406</td>
<td>799</td>
<td>640</td>
<td>862</td>
<td>707</td>
<td>525</td>
<td>516</td>
<td>463</td>
<td>546</td>
</tr>
<tr>
<td>2010</td>
<td>1902</td>
<td>2026</td>
<td>994</td>
<td>1009</td>
<td>1053</td>
<td>953</td>
<td>628</td>
<td>813</td>
<td>624</td>
<td>664</td>
</tr>
<tr>
<td>2011</td>
<td>2701</td>
<td>2373</td>
<td>1430</td>
<td>1446</td>
<td>1332</td>
<td>1217</td>
<td>1205</td>
<td>946</td>
<td>759</td>
<td>976</td>
</tr>
<tr>
<td>2012</td>
<td>1983</td>
<td>1845</td>
<td>1074</td>
<td>937</td>
<td>867</td>
<td>864</td>
<td>665</td>
<td>534</td>
<td>393</td>
<td>804</td>
</tr>
<tr>
<td>2013</td>
<td>1994</td>
<td>1782</td>
<td>968</td>
<td>1113</td>
<td>886</td>
<td>924</td>
<td>390</td>
<td>501</td>
<td>698</td>
<td>814</td>
</tr>
<tr>
<td>2014</td>
<td>1172</td>
<td>1109</td>
<td>758</td>
<td>850</td>
<td>501</td>
<td>464</td>
<td>209</td>
<td>328</td>
<td>712</td>
<td>637</td>
</tr>
</tbody>
</table>

As seen in Figure 1, there is a non-uniform distribution in the published tasting notes by year, which can cause bias when comparing the frequency of words from different years. To avoid this bias, the term frequency ratio for each year as shown in Equation 1 will be analyzed.

\[ TR(w_i, y_k) = \frac{TF(w_i, y_k)}{\sum_{j \in W} TF(w_j, y_k)} \] (1)

Where, the term frequency \( TF(w_i, y_k) \) of a word \( w_i \) occurring in a specific year \( y_k \) are divided by the sum of all the frequencies of words occurring in the same year.

The occurrence percentage for each year of the top 10 frequent adjectives are shown in Figure 2.

![Fig. 2. Top 10 frequently occurring adjectives by year.](image)

3. Diachronic Analysis of Antonyms by Frequency

In this section, we will examine the changes in the usage of the antonym pair thick and thin by analyzing the occurrence frequency of each word over time. Figure 3 shows the frequency of tasting notes that contain thick and thin from the year 1999 to 2014. It should be noted that for most years, wine tasting notes contain either thick or thin, and there are only a few exceptions that occur in the years 2002 and 2003 where both thick and thin occur in the same wine tasting note.

![Fig. 3. Distribution of the occurrence of adjectives thick and thin by year.](image)

As explained in the previous section, there is a non-uniform distribution in the number of published wine tasting notes year to year. The term frequency ratio of the occurrence of the words thick and thin for each year are shown in Figure 3. A rise in usage of thin between 2000 and 2002 occurs before the increased usage of thick between 2001 and 2003, suggesting that thin seems to lead in usage before thick.

Examining the relation of thick and thin with other adjectives over time can provide insight into the changes in use patterns in wine tasting notes. By measuring the dissimilarity of co-occurring words between thick and thin by year can show when the antonym pair are being used in similar patterns, and when they are not.

\[ \text{Dissimilarity}(a, b, y) = 1 - \frac{|W(a, y) \cap W(b, y)|}{|W(a, y) \cup W(b, y)|} \] (2)

Equation 2 is based on the Jaccard similarity distance and calculates the dissimilarity distance between two words \( a \) and \( b \) in the year \( y \), where \( W(a, y) \) represents the
words of the documents that contain the word \( a \) and were published in a wine tasting note in year \( y \).

Fig. 4. Jaccard dissimilarity of co-occurring adjectives of thick and thin by year.

The changes in the dissimilarity over the years of thick and thin by co-occurring adjectives is shown in Figure 4. In the year 1999 the dissimilarity between the pair is decreasing, until it reaches the lowest point in 2003. After this point, there is a steady increase in the dissimilarity distance, with a trough in 2009. This could be interpreted as showing the increasing dissimilarity in use patterns of thick and thin.

4. Diachronic Analysis of Antonyms by Mind Map

In this section we will investigate the differences in antonym use over time by observing changes in related words by mind map visualization.

4.1. Mind Map Generation

The method presented in this paper for generating mind maps is based on a system that was proposed in [2,3]. The root node of the mind map is specified as an initial search query from which the child nodes expand. Map expansion is limited to a set number of nodes or other features of the map. In this paper, the number of related words (nodes) in the graph was limited to the 10 top ranking words related to the root node. The document frequency of a word was used as the ranking to determine the expansion of the map. Firstly, simple mind maps were generated for each year of wine tasting notes containing either of the antonyms: thick or thin.

Fig. 5. A mind map of thin from 1999.

An example of the mind map of wine tasting notes from the year 1999 containing the word thin is shown in Figure 5. The root node that contains the word thin and the number of documents in which the word occurs: 9. Strong related word nodes expand out from the root node, in this case: short, flavorful, and fruitful are closely related to the root node thin.

Fig. 6. A mind map of thick from 1999.

Some of the mind maps generated contained common nodes, suggesting that there are common and distinct relations between antonym pairs. An example of this can be seen by comparing the mind map for thin in Figure 5 with the mind map for thick from the corresponding year in Figure 6. Both maps contain a fruitful node branching off from the root node. Other common nodes occur further away from the root node.

To identify and visualize the distinct and common characteristics of the thick/thin antonym pair, a merged
mind map of the two single word root node maps can be generated by combining common nodes.

Fig. 7. Merging mind maps of think and thin to show distinct and common nodes.

The merging process consolidates nodes that represent the same word in two or more different graphs, as seen in the example showing the merger of two sub maps in Figure 7. The consolidated node is connected by all the edges of the nodes it represents. The consolidated nodes are represented by a dashed outline and no fill color. Edges connecting consolidated nodes are also represented by dashed lines. In the example, the root node of the merged map contains both of the words from the two sub maps, and they are colored according to the representation of distinct nodes in the merged map, in this case: thick is red, and thin is blue. The distinct nodes in the merged map are filled and outlined accordingly: thick only nodes are pink filled and red outlined, thin only nodes are cyan filled and blue outlined. In the example merged map the fruitful node is dashed representing a common node, and tightness and palate are colored representing distinct nodes from thick and thin maps respectively.

Fig. 8. The merged mind map of think and thin from 1999.

In the full merged mind map of thick and thin from 1999 shown in Figure 8, there are 4 common nodes (flavorful, fruitful, oaked, and finishing), 5 distinct thick nodes, and 5 distinct thin nodes lying on the outside of the map. Two of the common node words (flavorful, fruitful) are from the top 10 most frequent adjectives that are listed in Table 1.

Fig. 9. Mind map of thick and thin from 2004.

The mind map of thick and thin from the year 2004 as seen in Figure 9 contains the highest ratio of common nodes out of all of the years, with only 2 distinct thick and thin nodes (heavy and richness, and acidity and lightness). The nodes heavy and lightness are also an antonym pair used to express weight, and this co-occurrence could suggest a relation with the thick and thin antonym pair. Table 2 shows that in two of the year’s co-occurrences in use were observed.

Table 2. The co-occurrence of the heavy/lightness antonym pair with thick/thin.

<table>
<thead>
<tr>
<th>Year</th>
<th>Heaviness (thick)</th>
<th>Lightness (thin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2004</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>2009</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2010</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2013</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Fig. 10. Mind map of thick and thin from 2014.

The mind map with the highest ratio of distinct nodes is that from 2014 as seen in Figure 10. It should be noted that there is a large imbalance in the number of wine tasting notes containing the word thick (30 documents) and thin (3 documents). Only the nodes fruitful and finishing are common between the antonym pair. As with the map from 2004, richness is associated with the word thick. This is also the case in the following years: 2000, 2002, 2003, 2004, 2005, 2006, 2012, 2013, and 2014.

The dissimilarity of mind maps from the same year can be analyzed to investigate how the relation between thick and thin has changed over time.

Fig. 11. Jaccard dissimilarity of the mind maps of thick and thin by year.

In figure 11, the Jaccard dissimilarity (see Equation 2) between the nodes of the mind maps of thick and thin are plotted over time to show the change in distinct and common nodes. As the mind maps are limited to 10 nodes each, only words with high relevance to thick and thin are analyzed. Compared with the dissimilarity analysis in the previous section, the dip in dissimilarity occurs a year later in 2004, and then increases steadily without a pronounced trough. It should be noted that the variance in dissimilarity is greater than the analysis on co-occurring adjectives.

5. Conclusion

In this paper, diachronic analysis of the use patterns of the antonym pair thick and thin in wine tasting notes from the year 1999 to 2014 was examined. In particular, the following was analyzed: the frequency of wine tasting notes in which the antonym pair occurred, the dissimilarity of co-occurring adjectives, and visualization of the dissimilarity by mind map. The analysis by co-occurring adjectives suggested an increase in the similarity of use patterns from 1999 to 2003, and then a steady increase in dissimilarity in the following year. The analysis by mind map also suggested a similar trend, however dissimilarity reached its lowest point a year later in 2004, suggesting a delay in the trend of highly related words.

In future work, we plan to investigate other antonym pairs to see if similar changes of use patterns occurred. Analysis into the changes in diversity of expressions is also required.

6. Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 15J04830.

References