Analysis of The Diachronic Relations of Adjective Antonym Pairs in Wine Tasting Notes

Brendan Flanagan¹, Carita Paradis², Kiyota Hashimoto³, and Sachio Hirokawa⁴

¹ Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan
² Centre for Languages and Literature, Lund University, Sweden
³ Faculty of Technology and Environment, Prince of Songkla University, Thailand
⁴ Research Institute for Information Technology, Kyushu University, Japan

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 patterns over time of antonym pairs in wine tasting notes. We examine the change in the use of thick and thin by analyzing words that co-occur in the same tasting note as thick or thin and generate visualizations for diachronic analysis.

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

1 INTRODUCTION

Adjectives play the role of expressing attributes and are frequently used in wine tasting notes to describe opposite sensory properties of the wines along various different meaning dimensions, often concatenated in sentences to describe complex characteristics, as seen in the following example from Paul Gregutt’s tasting note of the Waterbrook 2002 Syrah¹: “This hits the palate with a crisp, clean, thin seam of flavor, balanced but quite light”. In this paper, we investigate the change in antonym use in wine tasting notes over time. The diachronic analysis presented in this paper, targets the antonyms thick and thin which were also part of a study investigating the use of antonymic adjectives in the BNC in Paradis et al. [1]. It was reported that use patterns relating to the WIDTH dimension suggested a positive negative antonym relation, expressing opposing poles of the dimension. However, in the case of wine tasting notes, polar opposites 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. An example of this double negative relation can be seen in Steve Heimoff’s tasting note of the Woodbridge 2000 Merlot² expressing negative characteristics with the opposing poles of the WIDTH dimension: “Feels dry and thin on the palate, with a thick, unnaturally oaky taste”.

Firstly, we will investigate the changes in use of thick and thin by the number of documents relating to the antonym pair by year. As words surrounding a word indicate how and what context a word is used in, we will analyzed changes in the frequency and the dissimilarity of adjectives that co-occur with thick or thin in the same tasting note.

Secondly, we make use of visualizations of thick and thin to aid the analysis usage patterns and changes of closely related adjectives over time.

2 DATA COLLECTION

The target of the analysis is a corpus that consists of 91,010 wine tasting notes, or 255,966 sentences, that were collected from the Wine Enthusiast website³. The attributes of each wine, i.e. the date when the note was published or uploaded to the website, 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. As many wine tasting notes were published in both the Wine Enthusiast magazine and on the website at different times, the earliest publication date was selected for analysis. An overview of the non-uniform distribution of publication dates by year can be seen in Figure 1.

³ http://buyingguide.winemag.com/
The corpus was indexed to construct a term document matrix based search engine using GETA. The words were stemmed and a list of 2,488 adjectives that frequently occur in the wine tasting notes was used as a mask for word selection. The raw frequencies of the top 10 frequent terms by year are shown in Table 1.

Table 1. Occurrence frequency of top 10 terms.

<table>
<thead>
<tr>
<th>Year</th>
<th>flavorful</th>
<th>fruity</th>
<th>finishing</th>
<th>acidity</th>
<th>cherry</th>
<th>tannins</th>
<th>sweetness</th>
<th>dry</th>
<th>palate</th>
<th>ripeness</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>384</td>
<td>228</td>
<td>235</td>
<td>221</td>
<td>213</td>
<td>197</td>
</tr>
<tr>
<td>2001</td>
<td>1419</td>
<td>1236</td>
<td>1135</td>
<td>510</td>
<td>577</td>
<td>477</td>
<td>385</td>
<td>505</td>
<td>656</td>
<td>326</td>
</tr>
<tr>
<td>2002</td>
<td>2166</td>
<td>1904</td>
<td>1338</td>
<td>800</td>
<td>741</td>
<td>947</td>
<td>685</td>
<td>796</td>
<td>764</td>
<td>566</td>
</tr>
<tr>
<td>2003</td>
<td>2712</td>
<td>1728</td>
<td>1850</td>
<td>982</td>
<td>1024</td>
<td>1104</td>
<td>699</td>
<td>701</td>
<td>996</td>
<td>631</td>
</tr>
<tr>
<td>2004</td>
<td>3179</td>
<td>2253</td>
<td>1954</td>
<td>1129</td>
<td>1440</td>
<td>1266</td>
<td>1121</td>
<td>921</td>
<td>972</td>
<td>817</td>
</tr>
<tr>
<td>2005</td>
<td>3551</td>
<td>2321</td>
<td>2066</td>
<td>1246</td>
<td>1739</td>
<td>1255</td>
<td>1078</td>
<td>1371</td>
<td>1020</td>
<td>915</td>
</tr>
<tr>
<td>2006</td>
<td>4030</td>
<td>2846</td>
<td>2047</td>
<td>1662</td>
<td>1632</td>
<td>1303</td>
<td>1092</td>
<td>1410</td>
<td>1029</td>
<td>1214</td>
</tr>
<tr>
<td>2007</td>
<td>4172</td>
<td>2968</td>
<td>1868</td>
<td>1483</td>
<td>1874</td>
<td>1272</td>
<td>1241</td>
<td>1271</td>
<td>974</td>
<td>1084</td>
</tr>
<tr>
<td>2008</td>
<td>4634</td>
<td>3315</td>
<td>1731</td>
<td>1597</td>
<td>1862</td>
<td>1504</td>
<td>1411</td>
<td>1155</td>
<td>961</td>
<td>1188</td>
</tr>
<tr>
<td>2009</td>
<td>5242</td>
<td>3782</td>
<td>2130</td>
<td>1956</td>
<td>2406</td>
<td>1715</td>
<td>1962</td>
<td>1561</td>
<td>1321</td>
<td>1520</td>
</tr>
<tr>
<td>2010</td>
<td>4853</td>
<td>4312</td>
<td>2378</td>
<td>2499</td>
<td>2361</td>
<td>1966</td>
<td>1717</td>
<td>1788</td>
<td>1570</td>
<td>1318</td>
</tr>
<tr>
<td>2011</td>
<td>5350</td>
<td>4777</td>
<td>2850</td>
<td>2798</td>
<td>2593</td>
<td>2023</td>
<td>2611</td>
<td>1693</td>
<td>1742</td>
<td>1690</td>
</tr>
<tr>
<td>2012</td>
<td>4563</td>
<td>3714</td>
<td>2373</td>
<td>2366</td>
<td>1827</td>
<td>1682</td>
<td>1677</td>
<td>1166</td>
<td>834</td>
<td>1545</td>
</tr>
<tr>
<td>2013</td>
<td>4870</td>
<td>3823</td>
<td>2249</td>
<td>2556</td>
<td>1920</td>
<td>1960</td>
<td>1138</td>
<td>1233</td>
<td>1608</td>
<td>1794</td>
</tr>
<tr>
<td>2014</td>
<td>3029</td>
<td>2663</td>
<td>1704</td>
<td>2047</td>
<td>1109</td>
<td>1307</td>
<td>560</td>
<td>925</td>
<td>1493</td>
<td>1395</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.

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

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.

---

4 http://geta.ex.nii.ac.jp/
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 or introduce thick to be used.

To further investigate the reason behind the changes in usage patterns, we can examine the changes from different sensory viewpoints. In previous work, we have examined the automatic prediction of sensory sentiment characteristics described in wine tasting notes [2]. This method involved training and evaluating support vector machine (SVM) models by analyzing wine tasting note data that has been manually classified into four previously defined [3] sensory viewpoints: smell, taste, touch, and vision.

Fig. 4. Distribution of thick and thin by year from the viewpoints of smell, taste, touch, and vision.

We predicted the sensory viewpoints by applying the SVM models to each wine tasting note of the corpus. The distribution of thick and thin in relation to the sensory viewpoints is shown in Figure 4. The viewpoint of taste has a strong relation to the usage of thick and thin, with the usage of thin rising slightly before thick. It is interesting to see that the usage of only thin from the viewpoint of smell increased from the year 2001 to 2003 which is around the same time as the leading increase in the overall use of thin. This suggests that the use of thin from the smell viewpoint was a contributor to the rise in usage of thin 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 co-occurring words between thick and thin by year we can show when the antonym pair are being used in similar patterns, and when they are not.

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

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. 5. Dissimilarity of the co-occurring adjectives of thick and thin by year.

The changes in the dissimilarity over the years of thick and thin are shown in Figure 5. 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. This could be interpreted as showing the increasing dissimilarity in use patterns of thick and thin.

Fig. 6. Dissimilarity of the co-occurring adjectives of thick and thin by year from the viewpoints of smell, taste, touch, and vision.

Analyzing the dissimilarity distance from four sensory viewpoints, we can see that the dissimilarity between the co-occurring adjectives of thick and thin from the taste viewpoint has remained highly dissimilar without much change over the years. This would suggest that another viewpoint which was not the target of this analysis, and to a minor extent the smell viewpoint, are influential in the change of dissimilarity from around the year 2008 to 2012 as seen in Figure 6.
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 [4,5]. Firstly, a co-occurrence graph is generated of the words occurring in the documents returned by searching for the query words. The nodes represent the words within the documents, and the edges are weighted by the document frequency of co-occurring words. A mind map is generated by finding the minimum spanning tree of the co-occurrence graph. 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. Firstly, simple mind maps were generated for each year of wine tasting notes containing either of the antonyms: thick or thin.

Fig. 7. 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 7. The root node that contains the word thin and the number of wine tasting notes (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. 8. 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 7 with the mind map for thick from the corresponding year in Figure 8. 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 of both the thick and thin mind maps from the same year.

Fig. 9. Merging mind maps of thick 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 9. 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. 10.** The merged mind map of thick and thin from 1999.

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

**Fig. 11.** Mind map of thick and thin from 2002.

**Fig. 12.** Mind map of thick and thin from 2004.

The mind map of thick and thin from 2002 and 2004 as seen respectively in Figures 11 and 12 contain the highest ratio of common nodes out of all of the years, with only 2 distinct thick and thin nodes for each year: richness, blackness, dry, and very thick nodes, and little, dry, acidity and lightness thin nodes. It is interesting to see that the dry node which was associated with thin in 2002 has become associated with thick in 2004. This suggests that dry is not affected by changes in the width meaning dimension.

**Fig. 13.** Mind map of thick and thin from 2013.

The mind map with the highest ratio of distinct nodes is that from 2013 as seen in Figure 13. Only the nodes finishing, flavorful, and fruitful are common between the antonym pair. As with the map from 2002, richness and blackness are
associated with the word thick, and dry and little with the word thin.

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.

![Figure 14](image)

**Fig. 14.** Jaccard dissimilarity of the mind maps of thick and thin by year.

In figure 14, 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 earlier in 2002 and later in 2004, and then increases steadily with a pronounced trough in 2012.

5 CONCLUSION

In this paper, a diachronic analysis of the use patterns of the antonym pair thick and thin in wine tasting notes from the year 1999 to 2014 were examined. In particular, the following aspects were 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 earlier in 2002 and then later in 2004. Further analysis from sensory viewpoints revealed that there was a rise in the use of thick and thin from the taste viewpoint, and only in thin from the smell viewpoint. This suggests that the increased use of thin from the smell viewpoint was a contributor to the rise in usage of thin before thick.

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