

A Parallel Corpus-Driven Approach to Bilingual Oenology Term

Banks: How Culture Differences Influence Wine Tasting Terms

Vincent Xian Wang

Department of English

University of Macau

vxwang@um.edu.mo

Xi Chen

Department of Chinese and Bilingual Studies

The Hong Kong Polytechnic University

Department of English, University of Macau

yb77703@um.edu.mo

Songnan Quan

Department of Chinese and Bilingual

Studies

The Hong Kong Polytechnic University

songnan.quan@connect.polyu.hk

Chu-Ren Huang

Department of Chinese and Bilingual Studies

The Hong Kong Polytechnic University

The HK PolyU-PKU Research Centre on

Chinese Linguistics

churen.huang@polyu.edu.hk

Abstract

This paper describes the construction of an English-Chinese Parallel Corpus of wine reviews and elaborates on one of its applications – i.e. an E-C bilingual oenology term bank of wine tasting terms. The corpus is sourced from Decanter China, containing 1211 aligned wine reviews in both English and Chinese with 149,463 Chinese characters and 66,909 English words. It serves as a dataset for investigating cross-lingual and cross-cultural differences in describing the sensory properties of wines. Our log-likelihood tests revealed good candidates for the Chinese translations of the English words in wine reviews. One of the most challenging features of this domain-specific bilingual term bank is the dominant many-to-many nature of term

mapping. We focused on the one-to-many English-Chinese mapping relations of two major types: (a) the words without a single precise translation (e.g. “palate”) and (b) the words that are underspecified and involve ‘place-holder’ translation (e.g. “aroma”). Our study differs from previous bilingual CompuTerm studies by focusing on an area where cultural and sensory experiences favour many-to-many mappings instead of the default one-to-one mapping preferred in scientific and jurisprudential areas. This necessity for many-to-many mappings in turn challenges the basic design feature of many state-of-the-art automatic bilingual term-extraction approaches.

1 Introduction

The textual data of wines in the field of wine informatics are increasingly accessible to the

public through the Internet in this big data era. Wine reviews have been much criticised in terms of the use of metaphors for describing wine-tasting experience that often goes so free that it becomes “difficult to understand” (Demaeker, 2017, p.117). However, Croijmans, Hendrickx, Lefever, Majid and Van Den Bosch (2020) refuted this line of criticism by exhibiting high consistency in the use of wine terms in 76,410 wine reviews they gathered, which also effectively trained a classifier that automatically and rather accurately predicted the wine colour (red, white, or rose) and grape variety (n=30) in the wine reviews that were new to the classifier. Largely consistent with the results of Croijmans et al. (2020), López-Arroyo and Roberts (2014) found wine reviews used a limited repertoire of commonly-used words to convey specialised senses about wine tasting experience, while extending metaphorical applications of the words. The sensory experience of wine tasting was also studied in the frame of “motions” by Caballero (2017), who drew on cognitive linguistic research on motion events to examine the description of the aromas and flavours of wine “travelling” to sensory organs. Caballero gathered 12,000 wine notes in both English and Spanish and identified similarities and differences in the description of motions between the two languages.

Research in sensory sciences and informatics focuses on the extracting meaningful information from the wine reviews. For example, Valente, Bauer, Venter, Watson and Nieuwoudt (2018) introduced a new approach of using formal concept lattices to visualise the sensory attributes of Chenin blanc and Sauvignon blanc wines. Palmer and Chen (2018) employed a large-scale dataset of wine reviews to perform regression predictions on the grade and price of wines. In linguistic studies, wine reviews provide sensory descriptions for the research on language and cognition. Thus, comparative studies based on bilingual wine reviews, such as Chinese and

English, underline the issue of how sensory cognition is encoded across different languages.

Another possible research application is to build English-Chinese parallel corpus based on wine reviews for domain-specific machine translation or translation studies. Such corpora should follow established guidelines (e.g. Chang, 2004) in order to be sharable. Such a corpus is crucial as terms loaded with rich cultural tradition tend to be considered ‘untranslatable’. In computational term banks, it often leads to one-to-many (cf. Lim, 2018, 2019), many-to-one, or many-to-many mappings, although the mapping relations do not seem to have attracted in-depth research in computational terminology. In this paper, we propose a parallel corpus-driven approach to culturally bound bilingual terms discovery. In particular, we look at English-Chinese bilingual wine-tasting terminology. Since modern table wine culture and technology are mostly borrowed in the direction from the Western world to China, we focus on the one-to-many mapping of terms in E-C wine terminology. Therefore, this paper will address: a) how we are constructing the English-Chinese parallel corpus of wine reviews; b) the application of this parallel corpus in computational E-C oenology terminology.

Another important issue in computational terminology that our study will raise lies in the design criteria and evaluation metrics. The assumed ideal world criterion in bilingual term extraction is to achieve perfect one-to-one mapping. Previous studies on formal information dominant domains (sciences, technology, law etc.) worked well under this default criterion. However, what happens when the best terms in the target language vary in a wide range according to the context? Is there a better algorithm for this complex mapping issue?

Henri Bourgeois, Les Baronnes, Sancerre, Loire, France 2013



Silver
Decanter Asia Wine Awards

Country: France
Region: Loire
Sub-region: Sancerre
Grape(s): 100% Sauvignon Blanc
Producer: Henri Bourgeois
Alc: 12.5%
Distributor in China: ASC Fine Wines
Price: CNY 253
Note: Please consult local distributor for the retail price



Delightful green apple, grapefruit and lemon zest nose with pronounced mineral notes. More greenish tones and minerality on the palate, which is fresh, clean and firmly structured, with firm, lingering acidity. Good food wine.

Figure 1: A Wine Review at Decanter (English)

亨利博卢瓦庄园，桑榭尔男爵，桑塞尔，卢瓦河谷，法国 2013



银奖
Decanter 亚洲葡萄酒大赛

国家: 法国
产区: 卢瓦河谷
次级产区: 桑塞尔
葡萄: 100% 长相思
酿酒商: 亨利博卢瓦庄园 (Henri Bourgeois)
酒精度: 12.5%
中国经销商: ASC精品酒业
参考价格: 253元
*注释: 请以经销商实际销售价格为准



宜人的青苹果、柚子以及柠檬芬芳，矿物清香明显。入口是植物的芬芳和矿物味，清新、爽口而有架构，酸度稳固而持久。是一款非常适合佐餐的葡萄酒。

Figure 2: A Wine Review at Decanter (Chinese)

2 The Parallel Corpus

We describe our parallel corpus in terms of the source data (cf. 2.1) and our corpus construction method (cf. 2.2).

2.1 Source Data

Our data consists of bilingual reviews published on DecanterChina.com (醇鉴中国 chún jiàn zhōngguó), a website (www.decanterchina.com) presented by Decanter magazine, an international wine authority. Each wine review (酒评 jiǔ píng) is presented in both English and Chinese (cf. Figures 1 and 2). One of the present authors who is an accredited English-Chinese translator in China studied the bilingual reviews and confirmed that the Chinese reviews were the human translations of the English ones, ruling out the possibility that they were the outcome of automatic machine translation.

We crawled the data by means of “request” and “Beautiful Soup” of Python. The website contains data on thousands of wines and each wine has a separate introduction page. Each page displays the name, score, region, grape variety, producer, alcohol level, reference price, and reviews of the wine, which are the focus of this study. By targeting the English and the corresponding Chinese URLs with the use of the English/Chinese switch button on the top right corner of each webpage, we wrote the scripts to simulate the process of clicking on each wine page, and automatically collected the content in each page (Figure 3). We saved the content into data frame (Figure 4), and manually removed the noise and inequivalent pairs in the data.



Figure 5: Chinese Word Cloud



Figure 6: English Word Cloud

3.2 Key Words Translation

In order to examine the English-Chinese translation correspondence of certain words for the wine reviews, we used an alignment method to detect the word-pairs by the log-likelihood ratio estimation demonstrated in Rapp (1999). This calculation is based on the assumption according to the Distributional Hypothesis (Harris, 1954) – i.e. word meaning depends on its textual context. Hence, in the parallel corpus, if an English word and a Chinese word co-occur frequently in the parallel sentences, they are potentially good translation candidates for each other (Samuelsson & Volk, 2007).

| English Word | Acceptable Translation | Top 10 Scored Candidates |
|--------------|---|---|
| Palate | 风格 fēng gé (style); 口味 kǒu wèi (taste) | 风格 fēng gé (style); 现在 xiàn zài (now); 舌尖 shé jiān (tongue tip); 非常 fēi cháng (very); 很 hěn (very); 混酿 hùn niàng (blend); 这款 zhè kuǎn (this type); 带 dài (with); 酿造 niàng zào (brew) |
| Nose | 香气 xiāng qì (scent); 鼻腔 bí qiāng (nasal cavity) | 香气 xiāng qì (scent); 推荐人 tuī jiàn rén (recommender); 非常 fēi cháng (very); 款 kuǎn (type); 黑比诺 hēi bǐ nuò (Pinot Noir); 酒庄 jiǔ zhuāng (winery); 鼻腔 bí qiāng (nasal cavity); 潜力 qián lì (potential); 这款 zhè kuǎn (this type); 口味 kǒu wèi (taste) |
| Notes | nil | 几年 jǐ nián (several years); 不满 bú mǎn (dissatisfied); 木槿花 mù jǐn huā (hibiscus); 符合 fú hé (correspond); 水平 shuǐ píng (level); 月刊 yuè kān (monthly); 杂志 zá zhì (magazine); 明星 míng xīng (star); 坚固 jiān gù (firm); 甜椒 tián jiāo (sweet pepper) |
| Aromas | 芬芳 fēn fang (fragrance) | 强壮 qiáng zhuàng (strong); 黑色 hēi sè (black); 芬芳 fēn fang (fragrance); 黄油 huáng yóu (butter); 非常 fēi cháng (very); 完美 wán měi (perfect); 平衡 píng héng (balanced); 红果 hóng guǒ (red fruit); 十分 shí fēn (very); 草木 cǎo mù (vegetation) |

Table 1: The Word with its Top 10 Candidates

Since the corpus is already aligned in terms of English-Chinese pairs at the paragraph level, we can directly process the word alignment part. The word-level co-occurrence frequency was calculated and a statistical test for the log-likelihood ratio was launched. We desired to do the sentence alignment, but the way to segment a paragraph into the sentences is often different between Chinese and English. We excluded the stop words in both English and Chinese for the purpose of better decreasing the noise. We calculated the log-likelihood ratio for every possible pair of English-Chinese words in the corresponding wine review and sorted them according to the log-likelihood score.

As a result, we automatically extracted a bilingual lexicon from the parallel corpus on wine reviews. Four most frequently used words in the word clouds – i.e., “palate”, “nose”, “notes” and “aromas” – are presented in this section with their translation equivalents. Plural forms are used for “notes” and “aromas” here, since their singular forms occur at very low frequencies in our data. **Table 1** lists each word with its top 10 scored candidates, and we manually selected the ‘accepted translations’ from the top 10 based on their potential to serve as optional translations in wine reviews. The full list of our result, i.e. the English-Chinese bilingual oenology term bank (Chen et al., 2020) can be viewed at <https://drive.google.com/file/d/1LIDuU0euWKSzqWE1eUdzua25m6v3J4X/view?usp=sharing>.

First, the literal translation of “palate” is “腭” è in Chinese, which sounds odd for reviewing wines in Chinese culture as native speakers do not directly mention this sensory part to describe their wine-tasting experience. There are two acceptable translations of “palate” in the candidates, namely “风格” fēng gé (style) and “口味” kǒu wèi (taste), which are rather free renditions. The differences in cultural and sensory experiences between English and Chinese favour this one-to-many mapping instead of the default one-to-one mapping.

Second, the word “nose” can be either rendered very generally into “香气” xiāng qì (scent), or translated literally into “鼻腔” bí qiāng (nasal cavity), which preserves the semantic meaning of nose. Third, there are no acceptable translations of “notes” from the top candidates, and the log-likelihood scores of “notes” are not ideal. This points to a void in the Chinese lexis that corresponds to the meaning of “notes”. Finally, “aromas” tends to be rendered into “芬芳” fēn fang (fragrance), a very acceptable translation that beautifully conveys the meaning of aroma/s.

Our appliance of the log-likelihood leads to rather effective identification of good translation candidates for the English words in wine reviews, e.g., the translations for “palate”, “nose” and “aromas”. However, there were also cases in which not a single acceptable translation can be found – e.g., “notes” – which strongly suggest cross-linguistic and cross-cultural differences in word choices for wine reviews. The log-likelihood tests demonstrate that, based on the key words in wine reviews, translation candidates can be generated that are potentially useful for rendering the terms from English into Chinese (cf. ‘acceptable translations’ in Table 1). The translation candidates need to be manually selected to suit various oenological contexts though. Moreover, this method can be applied to the studies of the one-to-many and further many-to-many bilingual term extraction in the domain of oenology regarding the cross-lingual and cross-cultural differences. It can also involve translation studies that look into translation strategies – e.g. literal versus liberal, semantic versus communicative, or foreignised versus domesticated translation – in dealing with the specific texts of wine reviews and the manipulation of translators.

Based on the automatic extraction of mapping terms, two major types of mapping of words across the languages emerged. The first type of mapping pertains to the words that have no precise translation equivalent/s (e.g. “palate” and

“nose”), and therefore paraphrasing and other freer translation methods tend to be used. The second type involves ‘place-holder’ translation, while the multiple mappings are mostly dependent on the *modifiers* of the word in question to express different meanings. The second type exhibits two sub-types. The first sub-type is those that have null term mapping (e.g. “notes”). The term is so generic and flexible to collocate with a rich repertoire of modifiers that it is considered as a noun that is semantically bleached and is usually not translated, since the meaning is conveyed by the modifier/s of notes. The second sub-type (e.g. “aromas”) is still treated as semantically bleached but there is a corresponding ‘light noun’ – i.e., 芬芳 fēn fāng ‘fragrance’ – for direct (one-to-one) term mapping. Our subsequent task is to sort out a solution to automatically classify the three different types and represent these three different types of bilingual term mapping in a term bank.

4 Conclusion

In this paper, we introduced our English-Chinese parallel corpus of wine reviews and described our preliminary attempt for the extraction of bilingual oenology term bank. Our study showed that the log-likelihood approach we chose can deal with the many-to-many mapping challenge posed by the nature of ‘untranslatable’ terms. Yet it does require significant human intervention – i.e., manual selection of the useful translation candidates that suit different oenological contexts. On the other hand, the current corpus size is too small to support deep learning approaches. In the subsequent studies we will enlarge the corpus and also adopt a sensory domain based (rather than term-based) mapping to attempt more revealing findings.

Acknowledgements

We are thankful to the two anonymous reviewers of this article for their valuable comments and

suggestions. The authors would like to acknowledge the support of the research projects “A Comparative Study of Synaesthesia Use in Food Descriptions between Chinese and English” (G-SB1U) of The Hong Kong Polytechnic University and MYRG2018-00174-FAH of the University of Macau.

References

- Caballero, R. (2017). From the glass through the nose and the mouth: Motion in the description of sensory data about wine in English and Spanish. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 23(1), 66-88. <https://doi.org/10.1075/term.23.1.03cab>
- Chang, B. (2004). Chinese-English parallel corpus construction and its application. In H. Masuichi, T. Ohkuma, K. Ishikawa, Y. Harada, & K. Yoshimoto (Eds.), *Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation* (pp. 283-290). Tokyo, Japan: Logico-Linguistic Society of Japan.
- Chen, X., Quan, S., Wang, X., & Huang, C. (2020). An English-Chinese bilingual oenology term bank, ISLRN 851-636-882-375-0.
- Croijmans, I., Hendrickx, I., Lefever, E., Majid, A., & Van Den Bosch, A. (2020). Uncovering the language of wine experts. *Natural Language Engineering*, 26(5), 511-530. <https://doi.org/10.1017/S1351324919000500>
- Demaecker, C. (2017). Wine-tasting metaphors and their translation: A cognitive approach. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 23(1), 113-131. <https://doi.org/10.1075/term.23.1.05dem>
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3), 146-162. <https://doi.org/10.1080/00437956.1954.11659520>
- Lim, L. (2018). A corpus-based study of braised dishes in Chinese-English menus. In S. Politzer-Ahles, Y.-Y. Hsu, C.-R. Huang, & Y. Yao (Eds.),

Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation: 25th Joint Workshop on Linguistics and Language Processing (pp. 887-892). Association for Computational Linguistics.

Lim, L. (2019). Are TERRORISM and *kongbu zhuyi* translation equivalents? A corpus-based investigation of meaning, structure and alternative translations. In R. Otoguro, M. Komachi, & T. Ohkuma (Eds.), *Proceedings of the 33rd Pacific Asia Conference on Language, Information and Computation* (pp. 516-523). Association for Computational Linguistics.

López-Arroyo, B., & Roberts, R. P. (2014). English and Spanish descriptors in wine tasting terminology. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 20(1), 25-49. <https://doi.org/10.1075/term.20.1.02lop>

Palmer, J., & Chen, B. (2018). Wineinformatics: Regression on the grade and price of wines through their sensory attributes. *Fermentation*, 4(4), 84-93. <https://doi.org/10.3390/fermentation4040084>

Rapp, R. (1999). Automatic identification of word translations from unrelated English and German corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 519-526). College Park, Maryland: Association for Computational Linguistics. <https://doi.org/10.3115/1034678.1034756>

Samuelsson, Y., & Volk, M. (2007). Alignment tools for parallel treebanks. *GLDV Frühjahrstagung*. Tübingen, Germany: Zurich Open Repository and Archive, University of Zurich. <https://doi.org/10.5167/uzh-20448>

Valente, C.C., Bauer, F.F., Venter, F., Watson, B., & Nieuwoudt, H.H. (2018). Modelling the sensory space of varietal wines: Mining of large, unstructured text data and visualisation of style

patterns. *Scientific Reports*, 8(4987), 1-13. <https://doi.org/10.1038/s41598-018-23347-w>

Appendices

The 100 most frequently occurring wine-tasting words in both Chinese and English reviews.

Appendix A. The Chinese Top 100 Frequent Words

| Word | N of characters | Frequency | Weighted percentage (%) |
|------|-----------------|-----------|-------------------------|
| 风味 | 2 | 674 | 2.21 |
| 水果 | 2 | 504 | 1.65 |
| 香 | 1 | 468 | 1.53 |
| 酒 | 1 | 464 | 1.52 |
| 黑 | 1 | 413 | 1.35 |
| 款 | 1 | 387 | 1.27 |
| 非常 | 2 | 367 | 1.20 |
| 味 | 1 | 348 | 1.14 |
| 橡木 | 2 | 346 | 1.13 |
| 口感 | 2 | 331 | 1.08 |
| 单 | 1 | 324 | 1.06 |
| 浓郁 | 2 | 319 | 1.04 |
| 香气 | 2 | 300 | 0.98 |
| 味道 | 2 | 291 | 0.95 |
| 气息 | 2 | 290 | 0.95 |
| 成熟 | 2 | 277 | 0.91 |
| 芬芳 | 2 | 269 | 0.88 |
| 很 | 1 | 239 | 0.78 |
| 莓 | 1 | 235 | 0.77 |
| 酸度 | 2 | 229 | 0.75 |
| 中 | 1 | 219 | 0.72 |
| 余味 | 2 | 215 | 0.70 |
| 果 | 1 | 211 | 0.69 |
| 人 | 1 | 208 | 0.68 |
| 口味 | 2 | 208 | 0.68 |
| 葡萄酒 | 3 | 196 | 0.64 |
| 樱桃 | 2 | 190 | 0.62 |
| 般 | 1 | 181 | 0.59 |
| 黑色 | 2 | 179 | 0.59 |
| 风格 | 2 | 175 | 0.57 |
| 十分 | 2 | 164 | 0.54 |
| 感 | 1 | 162 | 0.53 |
| 香料 | 2 | 158 | 0.52 |
| 优雅 | 2 | 156 | 0.51 |
| 清爽 | 2 | 152 | 0.50 |
| 体 | 1 | 142 | 0.46 |

| | | | |
|-----|---|-----|------|
| 充满 | 2 | 142 | 0.46 |
| 复杂 | 2 | 140 | 0.46 |
| 气 | 1 | 133 | 0.44 |
| 甜美 | 2 | 133 | 0.44 |
| 清新 | 2 | 130 | 0.43 |
| 柔和 | 2 | 129 | 0.42 |
| 平衡 | 2 | 128 | 0.42 |
| 醋栗 | 2 | 128 | 0.42 |
| 好 | 1 | 124 | 0.41 |
| 李子 | 2 | 122 | 0.40 |
| 细腻 | 2 | 122 | 0.40 |
| 不 | 1 | 120 | 0.39 |
| 丝 | 1 | 119 | 0.39 |
| 咸 | 1 | 118 | 0.39 |
| 甜 | 1 | 117 | 0.38 |
| 滑 | 1 | 114 | 0.37 |
| 熏 | 1 | 113 | 0.37 |
| 辛 | 1 | 113 | 0.37 |
| 汁 | 1 | 110 | 0.36 |
| 令 | 1 | 106 | 0.35 |
| 分 | 1 | 106 | 0.35 |
| 烟 | 1 | 103 | 0.34 |
| 红 | 1 | 102 | 0.33 |
| 香草 | 2 | 99 | 0.32 |
| 迷人 | 2 | 98 | 0.32 |
| 纯净 | 2 | 97 | 0.32 |
| 淡淡 | 2 | 95 | 0.31 |
| 丰满 | 2 | 93 | 0.30 |
| 红色 | 2 | 93 | 0.30 |
| 美 | 1 | 93 | 0.30 |
| 带有 | 2 | 92 | 0.30 |
| 苹果 | 2 | 92 | 0.30 |
| 丰富 | 2 | 91 | 0.30 |
| 饱满 | 2 | 89 | 0.29 |
| 矿物 | 2 | 86 | 0.28 |
| 结构 | 2 | 86 | 0.28 |
| 活泼 | 2 | 85 | 0.28 |
| 出 | 1 | 82 | 0.27 |
| 巧克力 | 3 | 82 | 0.27 |
| 胡椒 | 2 | 81 | 0.27 |
| 还 | 1 | 81 | 0.27 |
| 陈年 | 2 | 81 | 0.27 |
| 柠檬 | 2 | 77 | 0.25 |
| 带来 | 2 | 76 | 0.25 |
| 更 | 1 | 76 | 0.25 |
| 紧 | 1 | 76 | 0.25 |
| 强劲 | 2 | 74 | 0.24 |

| | | | |
|----|---|----|------|
| 经典 | 2 | 74 | 0.24 |
| 具有 | 2 | 73 | 0.24 |
| 滋味 | 2 | 73 | 0.24 |
| 柑橘 | 2 | 72 | 0.24 |
| 氛 | 1 | 72 | 0.24 |
| 起来 | 2 | 71 | 0.23 |
| 会 | 1 | 70 | 0.23 |
| 口腔 | 2 | 70 | 0.23 |
| 绵长 | 2 | 70 | 0.23 |
| 饮 | 1 | 70 | 0.23 |
| 果香 | 2 | 69 | 0.23 |
| 回味 | 2 | 68 | 0.22 |
| 圆润 | 2 | 68 | 0.22 |
| 悠长 | 2 | 68 | 0.22 |
| 霞 | 1 | 68 | 0.22 |
| 完美 | 2 | 67 | 0.22 |
| 酿 | 1 | 66 | 0.22 |

Table 2: The Chinese Top 100 Frequent Words

Appendix B. The English Top 100 Frequent Words

| Word | N of characters | Freq | Weighted percentage (%) |
|----------|-----------------|------|-------------------------|
| palate | 6 | 692 | 3.01 |
| fruit | 5 | 588 | 2.56 |
| nose | 4 | 482 | 2.10 |
| aromas | 6 | 334 | 1.46 |
| wine | 4 | 311 | 1.35 |
| oak | 3 | 299 | 1.30 |
| finish | 6 | 294 | 1.28 |
| ripe | 4 | 266 | 1.16 |
| tannins | 7 | 264 | 1.15 |
| black | 5 | 232 | 1.01 |
| notes | 5 | 232 | 1.01 |
| acidity | 7 | 217 | 0.95 |
| fresh | 5 | 210 | 0.91 |
| sweet | 5 | 193 | 0.84 |
| red | 3 | 171 | 0.74 |
| well | 4 | 164 | 0.71 |
| fruits | 6 | 156 | 0.68 |
| cherry | 6 | 151 | 0.66 |
| flavours | 8 | 149 | 0.65 |
| spice | 5 | 144 | 0.63 |
| style | 5 | 135 | 0.59 |
| dark | 4 | 125 | 0.54 |
| juicy | 5 | 120 | 0.52 |
| long | 4 | 118 | 0.51 |

| | | | |
|--------------|----|-----|------|
| rich | 4 | 115 | 0.50 |
| elegant | 7 | 104 | 0.45 |
| cassis | 6 | 103 | 0.45 |
| savoury | 7 | 103 | 0.45 |
| vanilla | 7 | 99 | 0.43 |
| plum | 4 | 98 | 0.43 |
| fine | 4 | 97 | 0.42 |
| hints | 5 | 92 | 0.40 |
| lovely | 6 | 91 | 0.40 |
| full | 4 | 88 | 0.38 |
| bright | 6 | 82 | 0.36 |
| good | 4 | 82 | 0.36 |
| smoky | 5 | 82 | 0.36 |
| blackberry | 10 | 79 | 0.34 |
| character | 9 | 78 | 0.34 |
| apple | 5 | 75 | 0.33 |
| chocolate | 9 | 75 | 0.33 |
| clean | 5 | 75 | 0.33 |
| soft | 4 | 74 | 0.32 |
| texture | 7 | 74 | 0.32 |
| intense | 7 | 72 | 0.31 |
| floral | 6 | 71 | 0.31 |
| pepper | 6 | 70 | 0.30 |
| firm | 4 | 69 | 0.30 |
| touch | 5 | 69 | 0.30 |
| made | 4 | 68 | 0.30 |
| citrus | 6 | 67 | 0.29 |
| mineral | 7 | 66 | 0.29 |
| liquorice | 9 | 64 | 0.28 |
| spicy | 5 | 64 | 0.28 |
| balanced | 8 | 63 | 0.27 |
| bodied | 6 | 62 | 0.27 |
| complex | 7 | 61 | 0.27 |
| great | 5 | 60 | 0.26 |
| characters | 10 | 59 | 0.26 |
| shows | 5 | 59 | 0.26 |
| freshness | 9 | 57 | 0.25 |
| attractive | 10 | 56 | 0.24 |
| hint | 4 | 56 | 0.24 |
| peach | 5 | 56 | 0.24 |
| blackcurrant | 12 | 55 | 0.24 |
| green | 5 | 55 | 0.24 |
| structure | 9 | 54 | 0.24 |
| white | 5 | 54 | 0.24 |
| crisp | 5 | 53 | 0.23 |
| yet | 3 | 53 | 0.23 |
| creamy | 6 | 52 | 0.23 |

| | | | |
|---------------|----|----|------|
| dried | 5 | 52 | 0.23 |
| lemon | 5 | 51 | 0.22 |
| structured | 10 | 49 | 0.21 |
| berry | 5 | 48 | 0.21 |
| light | 5 | 48 | 0.21 |
| concentration | 13 | 47 | 0.20 |
| easy | 4 | 47 | 0.20 |
| pure | 4 | 47 | 0.20 |
| powerful | 8 | 46 | 0.20 |
| blueberry | 9 | 45 | 0.20 |
| herbs | 5 | 45 | 0.20 |
| lively | 6 | 45 | 0.20 |
| medium | 6 | 45 | 0.20 |
| cabernet | 8 | 44 | 0.19 |
| classic | 7 | 44 | 0.19 |
| concentrated | 12 | 44 | 0.19 |
| delicate | 8 | 44 | 0.19 |
| mouth | 5 | 44 | 0.19 |
| tannin | 6 | 44 | 0.19 |
| toasty | 6 | 44 | 0.19 |
| cherries | 8 | 43 | 0.19 |
| complexity | 10 | 43 | 0.19 |
| herbal | 6 | 43 | 0.19 |
| integrated | 10 | 43 | 0.19 |
| length | 6 | 43 | 0.19 |
| lime | 4 | 43 | 0.19 |
| spices | 6 | 43 | 0.19 |
| followed | 8 | 42 | 0.18 |
| dry | 3 | 41 | 0.18 |

Table 3: The English Top 100 Frequent Words