

This is a repository copy of *Uncovering the language of wine experts*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/171595/>

Version: Published Version

---

**Article:**

Croijmans, Ilja, Hendrickx, Iris, Lefever, Els et al. (2 more authors) (2020) Uncovering the language of wine experts. *Natural language engineering*. pp. 511-530. ISSN 1469-8110

<https://doi.org/10.1017/S1351324919000500>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

ARTICLE

# Uncovering the language of wine experts

Ilja Croijmans<sup>1,2\*</sup> , Iris Hendrickx<sup>2</sup>, Els Lefever<sup>3</sup>, Asifa Majid<sup>2,4,5,6</sup> and Antal Van Den Bosch<sup>2,7</sup>

<sup>1</sup>Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, the Netherlands, <sup>2</sup>Centre for Language Studies, Faculty of Arts, Radboud University, Nijmegen, the Netherlands, <sup>3</sup>Language and Translation Technology Team (LT3), Department of Translation, Interpreting and Communication, Ghent University, Ghent, Belgium, <sup>4</sup>Donders Institute for Brain, Cognition and Behavior, Nijmegen, the Netherlands, <sup>5</sup>Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, <sup>6</sup>Department of Psychology, University of York, Heslington, York, United Kingdom and <sup>7</sup>Meertens Institute, Royal Netherlands Academy of Arts and Sciences, Amsterdam, the Netherlands

\*Corresponding author. E-mail: [i.m.croijmans@uu.nl](mailto:i.m.croijmans@uu.nl)

(Received 20 August 2018; revised 22 July 2019; accepted 25 July 2019; first published online 23 September 2019)

## Abstract

Talking about odors and flavors is difficult for most people, yet experts appear to be able to convey critical information about wines in their reviews. This seems to be a contradiction, and wine expert descriptions are frequently received with criticism. Here, we propose a method for probing the language of wine reviews, and thus offer a means to enhance current vocabularies, and as a by-product question the general assumption that wine reviews are gibberish. By means of two different quantitative analyses—support vector machines for classification and Termhood analysis—on a corpus of online wine reviews, we tested whether wine reviews are written in a consistent manner, and thus may be considered informative; and whether reviews feature domain-specific language. First, a classification paradigm was trained on wine reviews from one set of authors for which the color, grape variety, and origin of a wine were known, and subsequently tested on data from a new author. This analysis revealed that, regardless of individual differences in vocabulary preferences, color and grape variety were predicted with high accuracy. Second, using Termhood as a measure of how words are used in wine reviews in a domain-specific manner compared to other genres in English, a list of 146 wine-specific terms was uncovered. These words were compared to existing lists of wine vocabulary that are currently used to train experts. Some overlap was observed, but there were also gaps revealed in the extant lists, suggesting these lists could be improved by our automatic analysis.

**Keywords:** Corpus linguistics; Information extraction; Semantics; Statistical methods; Wine expertise

## 1. Introduction

### 1.1 *The language of wine experts*

Everyone begins as a novice, but through training and practice, one can obtain comprehensive and authoritative knowledge (i.e., epistemic expertise), and become more skilled in performing certain acts (i.e., performative expertise), and as such, become an expert (Weinstein 1993; Caley *et al.* 2014). Studies of expertise range from classic work on chess masters (de Groot 1946) and chicken sexers (Biederman and Shiffrar 1987) to studies of professional musicians (Mitchell and MacDonald 2011), sailors (Pluijms *et al.* 2015), and Japanese incense masters (Fujii *et al.* 2007).

Chess expertise, with its standardized levels of performance (i.e., the Elo rating system, named after its creator Arpad Elo), has been studied extensively and has informed models for how expertise is acquired more generally (de Groot 1946; De Groot 1978; De Groot *et al.* 1996; Ericsson *et al.* 2018). In domains other than chess, experts have also been found to perform better on various

cognitive tasks. For example, expert radiologists are better at detecting low-contrast features in X-ray images (Sowden, Davies and Roling 2000; Ericsson *et al.* 2007). Likewise, expert musicians are able to identify relationships between tones, that is, relative pitch (Levitin and Rogers 2005), imagine musical pieces from musical notations (Brodsky *et al.* 2003), and recall musical pieces more consistently than novices (Halpern and Bower 1982).

Similar effects have been shown with respect to linguistic skills too. When computer experts and novices are asked to describe pictures of complex visual scenes containing computer or other electronic equipment, experts' descriptions contain more references to salient details about the computer equipment (Humphrey and Underwood 2011). In line with this, when bird and dog experts are asked to list features of birds and dogs, they list more specific features for stimuli in their domain of expertise (Tanaka and Taylor 1991), suggesting more detailed conceptual representations.

The few studies investigating expertise effects on language have primarily done so using stimuli from the auditory or visual domain, but rarely investigated smells. It has been claimed that smell might be “ineffable” (Levinson and Majid 2014), without dedicated vocabulary across the world's languages (Sperber 1975), and experimental studies suggest odors are difficult to name (Cain 1979; Engen 1987; Cain *et al.* 1998); for reviews, see (Yeshurun and Sobel 2010; Olofsson and Gottfried 2015). The basis of this limitation is sought in our biological infrastructure: the words for smells may simply be inaccessible (Rivlin and Gravelle 1984; Lorig 1999), or the odor percept information may arrive relatively unprocessed to cortical areas responsible for language (Olofsson and Gottfried 2015). However, recent studies question whether poor odor naming is truly universal, showing that some populations are more eloquent when it comes to smells (Burenhult and Majid 2011; Majid and Burenhult 2014; Wnuk and Majid 2014; Majid 2015; Croijmans and Majid 2016; O'Meara and Majid 2016; San Roque *et al.* 2015; De Valk *et al.* 2017; Majid *et al.* 2018). Instead of universality, cultural factors, for example, subsistence, can shape how eloquent one is when it comes to naming smells (Majid and Kruspe 2018). Together, this research suggests that both across cultures and within sub-cultures, specific experience may be an important factor in how smells are talked about.

Wine experts—such as vinologists, sommeliers, and wine journalists—are an interesting group to study in this regard. Wine experts work with wines on a daily basis and communicate about the smell and flavor of wine in conversations among themselves and with consumers during wine tastings, as well as when writing tasting notes (Herdenstam *et al.* 2009). In these tasting notes and reviews, wines are often described following a set script: first the appearance of the wine is described, followed by smell (i.e., orthonasal olfaction), then flavor, and finally mouthfeel (Paradis and Eeg-Olofsson 2013). Flavor is defined as the combination of taste, smell (i.e., retronasal olfaction), trigeminal activation, and tactile sensation in the mouth (Auvray and Spence 2008; Smith 2012; Spence 2015b; Boesveldt and de Graaf 2017), with olfaction playing the major role in the experience of flavor (Shepherd 2006; Spence 2015a). This underscores the importance of both olfaction and language in wine expertise (Royet *et al.* 2013).

## 1.2 Wine reviews: *Intentional gibberish or consistent prose?*

Even though language features heavily in their expertise, wine experts often complain of its lack. In the words of wine journalist Malcolm Gluck (2003):

We wine writers are the worst qualified of critical experts. This is largely, though not exclusively, because we are the most poorly equipped. The most important tool at our disposal is inadequate for the job. That tool is the English language. (Gluck 2003, p. 107)

Scholars have suggested wine reviews are useless at informing readers about the flavor of wines. For example, Quandt (2007) claims “the wine trade is intrinsically bullshit-prone and therefore

attracts bullshit artists” (Quandt 2007, p. 135). Similarly, Shesgreen (2003) states wine reviews are “mystifying babble used by writers whose prose is deeply disconnected from the beverage they pretend to describe” (Shesgreen 2003, p. 1). Finally, Silverstein (2006) has suggested that “wine-talk” says as much about the speaker, as it does about the wine. In line with this line of critique, an experimental study by Lawless (1984) found that descriptions written by wine experts were highly idiosyncratic, with most terms only used once by one expert (Lawless 1984).

This raises the question of whether experts really can describe smells and flavors in a consistent way. The previous literature does not provide a satisfactory answer. Solomon (1990) examined whether experts and novices could match wine reviews produced by other experts and novices to the original wines. Novices were no better than chance at matching descriptions from experts, suggesting reviews produced by those experts were not particularly informative. In contrast, Gawel (1997), using a similar paradigm, found both experts and novices correctly matched descriptions to wines significantly above chance when the descriptions were produced by experts.

A previous computational linguistics study also suggests wine reviews can be considered consistent (Hendrickx *et al.* 2016). Hendrickx *et al.* (2016) used the text of wine reviews written by experts to predict the color, grape variety, origin, and price of wines. They found that there was enough consistency in terminology used by wine experts that information in the review text could distinguish classes of wine. Although promising, this study had some drawbacks. For example, wine reviews differ by authors’ personal vocabulary preferences. Just like other writers (Zheng *et al.* 2006; Juola 2008); Kestemont *et al.* (2012a,b), wine experts have been found to differ in their idiolectal use of lexical and syntactic features (Brochet and Dubourdieu 2001); Sauvageot, Urdapilleta and Peyron (2006); (Parr *et al.* 2011). The corpus used by Hendrickx *et al.* (2016) contained reviews from several authors but not an equal number of reviews per author. In fact, one author contributed almost 20 times as many reviews as the author who wrote the least. This skewing means that a single author consistent in their own description may have inflated the apparent consistency across writers. As such, a more rigorous test is required to establish whether wine reviews are consistent, in contrast to the view outlined above. We develop and evaluate this proposal here.

### 1.3 Domain-specific language in wine reviews

When wine experts talk about wines, they convey the smell and flavor of wine using various strategies. Wine experts famously employ metaphors in wine descriptions (Suárez Toste 2007; Caballero and Suárez-Toste 2010; Paradis and Eeg-Olofsson 2013). In addition, they use a set of conventionalized descriptors. Croijmans and Majid (2016) found that wine experts use more source descriptions (e.g., *red fruit*, *vanilla*) for describing the smell and flavor of wine than novices, whereas novices used more evaluative terms (e.g., *nice*, *lovely*). Other studies suggest experts use more specific, concrete words; for example, they say *blackberry* instead of *fruit* (Lawless 1984; Solomon 1990; Gawel 1997). Experts are also said to use more words for grape type and terroir (i.e., the origin of a wine) than novices (Parr *et al.* 2011).

To help budding wine enthusiasts to learn about wine, and enable description of wine flavors, expert tools have been developed that display lists of words deemed helpful. These words are often hierarchically ordered by their specificity and category—so-called “wine wheels” (Noble *et al.* 1984; Lehrer 2009). Various wine vocabularies exist, ranging from the wheel first created by Noble and colleagues (Noble *et al.* 1984, 1987) to wheels specific for red, white, or fortified wine, or wine from specific countries (e.g., wines from Germany). Other lists zoom in on specific aspects of wine flavor, such as the mouthfeel wheel (Gawel, Oberholster and Francis 2000), or are composed by a specific author (Parker 2017). The wide range of wine vocabulary lists again suggests wine vocabulary is diverse, which raises the question of whether these lists capture the terminology employed by a broad range of experts in actual language use. If this is not the case, learning to become an expert using these lists might not be as effective as it could be. Therefore, in this

study, we examined the specific terminology employed in wine reviews, so that the outcome of our method could potentially enhance or lead to the adaptation of current expert tools.

#### 1.4 The present study

To test whether reviews provide consistent information about wines (i.e., they are not bullshit; cf. Quandt 2007), an automatic classifier was trained with reviews from one set of authors and then used to predict properties of a new set of wine reviews written by a different author. By taking training data from one set of authors and testing on data from a different author, it was possible to establish description consistency between authors.

In a second set of analyses, we employed Termhood analysis to establish what words were used by each author to describe wine compared to a standard corpus of English. The words ranked high on Termhood were then analyzed using principal component analysis (PCA) to further establish whether there was consistency in language use across authors. Finally, the set of words ranked highest on Termhood was compared to previously established word lists of wine vocabulary (Noble *et al.* 1984; Lehrer 2009; Lenoir 2011; Parker 2017), to explore what similarities and differences exist in wine language tools.

## 2. Predicting wine properties and author differences

### 2.1 Methods

#### 2.1.1 Corpus description

A total of 76,410 wine reviews were collected from the internet<sup>a</sup> and assembled into a corpus. According to the source website:

All tastings reported in the Buying Guide are performed blind. Typically, products are tasted in peer-group flights of 5–8 samples. Reviewers may know general information about a flight to provide context—vintage, variety or appellation—but never the producer or retail price of any given selection.

The corpus contained structured information, that is, metadata about each wine, that is, price, designation, grape variety, appellation region, producer, alcohol content, production size, bottle size, category, importer, and when the wine was reviewed. In addition, each entry also contained a rating (range 80–100)<sup>b</sup> and a compact review describing the wine (on average approximately 40 words per review).

As prediction scores are affected by the amount of data used as input, only authors who had reviewed more than 1000 wines were considered, resulting in a selection of 13 authors. The contributions of these authors were not evenly distributed, with some authors producing around 1000–2000 reviews while the most prolific reviewer (later referred to as Author 1) contributed around 19,000 reviews, that is, 26% of all the reviews in the corpus. Altogether, a corpus of 73,329 reviews for these 13 authors was compiled. We did not further downsample or weigh the uneven contributions of the authors in order to stay as close as possible to the real distribution as found “in the wild.”

#### 2.1.2 Data preprocessing

The review texts were first preprocessed by means of the Stanford CoreNLP toolkit (Manning *et al.* 2014) which added linguistic information to reviews. The following steps were taken:

<sup>a</sup>The reviews were collected from the website <http://www.winemag.com>

<sup>b</sup>The scale theoretically ranges from 0 to 100 but wines with scores less than 80 typically do not receive reviews.

**Table 1.** Example output of preprocessing for the classification analysis

| Input    | Lemma    | Grammatical category | Kept for classification analysis |
|----------|----------|----------------------|----------------------------------|
| The      | the      | determiner           | –                                |
| wine     | wine     | noun                 | wine                             |
| has      | have     | auxiliary verb       | –                                |
| an       | an       | determiner           | –                                |
| easy     | easy     | adjective            | easy                             |
| approach | approach | noun                 | approach                         |
| .        | .        | .                    | –                                |

- (1) Tokenization: split review text into tokens (i.e., words, punctuation, numbers, etc.)
- (2) Part-of-speech tagging: assign grammatical category to tokens (e.g., noun, verb, etc.)
- (3) Lemmatization: provide lemma for tokens (nouns, adjectives: singular form; verbs: infinitive form)

Table 1 shows an example for the review sentence: “The wine has an easy approach.” We used this linguistic information to reduce texts to a vector of content words. To this end, we selected all single words with the grammatical labels noun, verb, adjective, and adverb and used their low-erased lemma for the classification studies. For this example sentence, the terms “*wine*,” “*easy*,” and “*approach*” were included in the vector representation of this sentence. For the experiments, these vectors were transformed into a new vector containing binary features, indicating whether a content word is present (feature value: “1”) or absent (value: “0”) in the review text.

### 2.1.3 Classification tasks

Several different classification tasks were executed based on the metadata available: color, grape variety, and origins. For **color**, three class labels were distinguished: red, white, and rosé. Wine reviews with the metadata color “unknown” ( $n = 5105$ ) were excluded. The words *red*, *white*, *rose*, and any variants thereof (i.e., *red*, *reds*, *white*, *whites*, *rosè*, *rosé*, *rose*) were removed from the wine reviews, so the classification could not be based on these terms.

For **grape variety**, only wines produced from a single grape were considered; blends were excluded. Different names used for the same grape in the metadata were normalized to get a consistent label for grape varieties for the classifier (but not in reviews): for example, *pinot gris* and *pinot grigio* were normalized into *pinot gris*. Only those grape class labels for which there were at least 200 reviews were included, resulting in 30 class labels: *aglianico*, *albarino*, *barbera*, *cabernet franc*, *cabernet sauvignon*, *carmenère*, *chardonnay*, *chenin blanc*, *gamay*, *glera*, *grenache*, *gruner veltliner*, *malbec*, *merlot*, *muscat*, *nebbiolo*, *nero d’avola*, *petite sirah*, *pinot blanc*, *pinot gris*, *pinot noir*, *riesling*, *sangiovese*, *sauvignon blanc*, *syrah*, *tempranillo*, *torrontes*, *traminer*, *viognier*, and *zinfandel*.

The wines had diverse origins: that is, 47 countries and over 1400 regions. We investigated the classification of **origin** using a coarse distinction, namely old versus new world (Remaud and Couderc 2006; Banks and Overton 2010). Broadly speaking, old world wines (e.g., France, Germany, Spain, and Italy) are “tradition-driven”: producers aim to make a high-quality product that can age well using traditional methods and terroir standards. In contrast, new world wines (e.g., USA, New Zealand, and Australia) are often produced with the latest production methods, and producers aim to make a good “consumer-driven” product in reasonable volumes that are valued by diverse consumer markets. Countries were categorized and labeled with the class labels



**Table 2.** List of countries considered new world, old world, or that were excluded from the origin task

| Class label | Countries  |
|-------------|--|
| New World   | Argentina, Australia, Brazil, Canada, Chile, China, Mexico, New Zealand, Peru, South Africa, South Korea, USA, Uruguay   |
| Old world   | Austria, France, Germany, Greece, Italy, Luxembourg, Montenegro, Portugal, Spain, Switzerland  |
| Excluded    | Bosnia Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Egypt, Georgia, Hungary, India, Israel, Japan, Lebanon, Lithuania, Macedonia, Moldova, Morocco, Romania, Serbia, Slovakia, Slovenia, Tunisia, Turkey, Ukraine |

“old” and “new” world based on country of origin. Reviews where the status was ambiguous (e.g., Eastern European countries) were excluded from consideration (see Table 2).

The machine learning classifier used in this study was support vector machines (SVMs) which performs particularly well on text classification tasks (Joachims 2002). The implementation LIBSVM (Library for Support Vector Machines) of SVM was used with the linear kernel (Chang and Lin 2011). The hyperparameter *C*, which controls the trade-off between setting a larger margin and lowering the misclassification rate, was optimized by means of a grid search on 10000 randomly selected instances of the training set. This optimization resulted in a cost value (*C*) of 0.125 for all three classification tasks.

The corpus contained reviews by 13 different wine experts. For all experiments per classification task (color, grape variety, and origin), we performed 13 leave-one-author-out iterations, such that a training sample contained reviews for 12 authors, after which the classifier was tested on the reviews of the remaining author. This setup was repeated 13 times. We calculated precision, recall, and F-score (Van Rijsbergen 1979) to measure the performance on all class labels.

For each class label, precision was computed as:

$$Precision = \frac{(Number\ of\ correctly\ predicted\ class\ labels)}{(Total\ number\ of\ predicted\ class\ labels)} \tag{1}$$

Recall was computed as:

$$Recall = \frac{(Number\ of\ correctly\ predicted\ class\ labels)}{(Number\ of\ gold\ standard\ classlabels)} \tag{2}$$

Finally, F-scores were calculated as follows:

$$F = (2 \times \frac{(Precision * Recall)}{(Precision + Recall)}) \tag{3}$$

To arrive at an overall F-score per author, we aggregate the precision and recall scores per class label for all reviews for the held-out author, that is, computed at the level of individual classifications. In the same way, by aggregating all individual classification over all held-out authors, we compute the overall F-score for one experiment (color, grape variety, or origin).

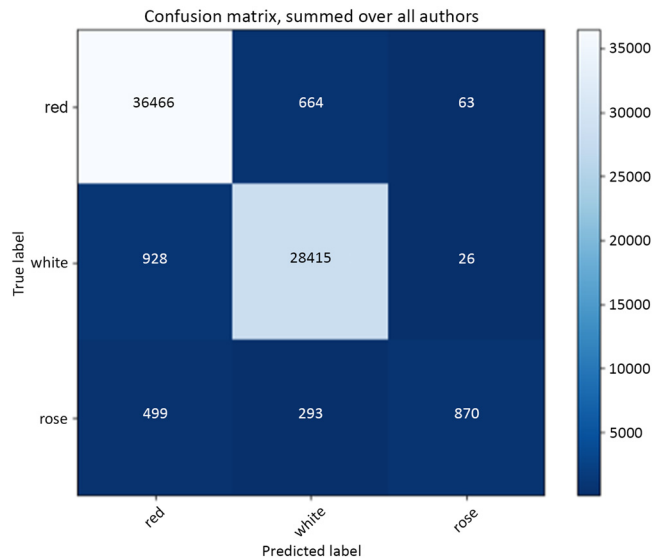
To estimate the predictive value of a classifier, F-scores were compared to an F-score majority baseline resulting from guessing the most frequent class label in that task (i.e., color, grape variety, and origin). For example, in the color classification task (with a total of 68224 reviews), the class label red was most frequent, with 36,466 reviews describing a red wine. If the classifier categorized each review as “red,” it would achieve a baseline F-score of 65.8%. Achieving this or a lower F-score would indicate that reviews are not consistently written.

**2.2 Results**

Table 3 lists the overall F-scores of the three classification tasks compared to their respective baselines. While the color task and grape variety task perform well above their respective majority

**Table 3.** Overall F-scores on each of the three different classification tasks across the 13 authors

| Task          | Number of reviews | F-score | Baseline |
|---------------|-------------------|---------|----------|
| Color         | 68,224            | 95.4    | 65.8     |
| Grape variety | 48,760            | 57.1    | 14.0     |
| Origin        | 72,925            | 61.5    | 56.0     |



**Figure 1.** Confusion matrix for the wine color classification task. Color shading indicates the relative number of individual classifications per cell with more classifications indicated by lighter cells.

baseline scores, the origin task appears to be the harder of the three. We discuss the performance on each task in more detail next and zoom in on the confusion matrices for each of the classification tasks.

For the color task, the classifier was able to predict the color of red and white wines well, suggesting there was consistency in the reviews produced by different authors. Recall that all reviews in the training set were written by authors different from those in the test set, due to the leave-one-author-out setup. Table 4 lists the F-scores per author. F-scores for the red and white colors are high and close to the mean for all authors. For rosé, we observe markedly lower F-scores for all authors and a somewhat larger variation in scores. To provide further insight on the distribution of the classifications over class labels against the true labels, we present the confusion matrix for the color classification task in Figure 1. Note this confusion matrix counts labels aggregated over all authors.

As Figure 1 shows, the two class labels red and white dominate both the true distribution and the predicted distribution. The relatively problematic third minority class label, rosé, is predicted correctly only 870 times. Rosé is most often misclassified as red wine (499 cases, that is, 63% of the misclassifications), and less often as white (293 cases, 37%). Therefore, the recall for the rosé class label is only 52. We compute the recall by dividing the number of correct predictions, 870 by the total number of times that it should have been predicted, that is, the sum of cell counts of the bottom row,  $499 + 293 + 870 = 1662$ . To calculate precision for rosé, 870 correct classifications were divided by the total of  $(63 + 26 + 870 = 959)$  rosé class label predictions, which amounts to a precision of 91. The F-score for the rosé class label is 66. Precision and recall scores for the other two color class labels are markedly higher; for red, precision is 96 and recall is 98, and for white, both precision and recall are 96.



**Table 4.** Number of reviews and F-scores per author, per class label and aggregated over the three class labels of the wine color task

| Author    | Red wine          |         | White wine        |         | Rosé wine         |         | Overall |
|-----------|-------------------|---------|-------------------|---------|-------------------|---------|---------|
|           | Number of Reviews | F-score | Number of Reviews | F-score | Number of Reviews | F-score |         |
| Author 1  | 13050             | 98.1    | 5364              | 96.8    | 302               | 44.3    | 97.1    |
| Author 2  | 5569              | 92.2    | 4108              | 90.5    | 455               | 26.1    | 89.8    |
| Author 3  | 6882              | 97.8    | 2261              | 94.6    | 231               | 33.1    | 95.9    |
| Author 4  | 5287              | 97.1    | 2098              | 94.5    | 157               | 52.7    | 95.6    |
| Author 5  | 4127              | 97.0    | 1648              | 93.7    | 61                | 44.6    | 95.5    |
| Author 6  | 2831              | 96.4    | 2403              | 96.7    | 189               | 40.3    | 95.3    |
| Author 7  | 2578              | 98.3    | 1011              | 95.8    | 39                | 36.9    | 97.1    |
| Author 8  | 1147              | 99.4    | 476               | 99.1    | 14                | 78.6    | 99.2    |
| Author 9  | 1042              | 98.2    | 400               | 97.7    | 44                | 60.9    | 97.2    |
| Author 10 | 828               | 97.2    | 395               | 96.0    | 23                | 29.6    | 96.1    |
| Author 11 | 345               | 94.7    | 754               | 98.3    | 42                | 51.6    | 95.9    |
| Author 12 | 621               | 97.8    | 412               | 97.2    | 33                | 61.2    | 96.7    |
| Author 13 | 572               | 96.9    | 353               | 98.5    | 72                | 61.0    | 95.6    |
| Total     | 44879             | 96.9    | 21683             | 95.5    | 1662              | 41.0    | 95.4    |

For grape variety, a random classifier would, on average, score no higher than 3% accuracy, and the majority baseline classifier (i.e., always predicts the most frequent class label; in this case *chardonnay*) would not achieve higher than 14% accuracy. The classifier, in fact, performed well above this baseline for all authors, as shown in Table 5.

Figure 2 displays the confusion matrix for the grape variety experiment, summed over all authors. Correct classifications are again visible as lighter cells populating the diagonal. The figure shows that classification of the *chardonnay* and *pinot noir* varieties was particularly good. For example, *pinot noir* is correctly recognized 5171 times out of 6667 instances. If we look at grape types that are most often confused with *pinot noir*, we observe that these are all red grape types. The label *pinot noir* is confused with *cabernet sauvignon* (486 times), *syrah* (393 times), *sangiovese* (182 times), and *chardonnay* (only 99 times). *Chardonnay* is also the most frequent white grape type label in the whole set.

As an example of a class label with an intermediate frequency, we examined the *riesling* label. *Riesling* is classified correctly in 1515 out of 2646 reviews with this label. It is most often confused with *chardonnay* (554 times), followed by *sauvignon blanc* (199), and *pinot gris* (186)—all white grape types—and only 44 times with *pinot noir* (a high frequent class, but red grape type).

For the origin classification, the majority baseline would predict the most frequent label to be new world (i.e., an F-score of 56.0 as shown in Table 3). The results showed an overall F-score of 61.5% which is a mere 5.5% above the majority baseline (Table 6). This could be because authors often show specialization for wines from a specific part of the world. For example, Authors 4 and 8 only reviewed wines from the old world, while Author 9 only reviewed wines from the new world. The classifier was able to predict origin above baseline for some authors (e.g., Authors 1, 9, and 10) but not for other authors (e.g., Author 2). The overall low score relative to baseline is unexpected and suggests authors do not describe new world wines distinctly from old world wines.

**Table 5.** Number of reviews and F-score for each author for grape variety

| Author    | Number of reviews | F-score |
|-----------|-------------------|---------|
| Author 1  | 16268             | 68.8    |
| Author 2  | 4796              | 41.7    |
| Author 3  | 6457              | 35.2    |
| Author 4  | 4547              | 45.6    |
| Author 5  | 4661              | 59.6    |
| Author 6  | 4015              | 64.1    |
| Author 7  | 3010              | 53.8    |
| Author 8  | 759               | 48.9    |
| Author 9  | 1193              | 77.0    |
| Author 10 | 896               | 57.1    |
| Author 11 | 986               | 78.9    |
| Author 12 | 618               | 69.4    |
| Author 13 | 554               | 59.4    |
| Total     | 48760             | 57.1    |

Figure 3 reflects the observations made on the basis of Table 5 regarding the performance of the classifier for distinguishing old world and new world wines. When summed over all authors, the classifier misclassifies more than half of the cases (18916) of old world wines (out of a total of 32082) as new world wines, yielding a recall of only 41 and a precision of 59.

Overall, these results show it is possible to predict color and grape variety from wine reviews, even when the classifier is trained using data written by authors other than that used as test data. This suggests experts describe wines in a predictable manner, and are consistent with other experts. Personal vocabulary idiosyncrasies do not seem to adversely impact the ability to assign wines into clearly meaningful classes.

### 3. The use of domain-specific language in wine reviews

#### 3.1 Methods

The earlier analysis shows that authors are using consistent terminology to distinguish classes of wine, but it does not reveal the terminology itself. We therefore sought to understand better what, if any, domain-specific vocabulary is used distinctively by wine experts, and if any such vocabulary was used consistently across authors. In order to do this, we examined Termhood, a key concept of terminology research, which refers to the degree that a linguistic unit is related to (or more straightforwardly represents) domain-specific concepts (Kageura and Umino 1996). The intuition is that Termhood expresses how much more frequent a word or word  $n$ -gram (i.e., a consecutive sequence of  $n$  words) is in the domain-specific wine corpus compared to a general corpus of English. The higher the Termhood value of a specific word, the more specialized that word is in comparison to its use in standard language use.

To extract terms belonging to domain-specific wine vocabulary, we used TExSIS, Terminology Extraction for Semantic Interoperability and Standardization (Macken, Lefever and Hoste 2013), a hybrid terminology extraction pipeline that combines linguistic and statistical information to extract domain-specific terms, that is, word  $n$ -grams, from a text corpus.

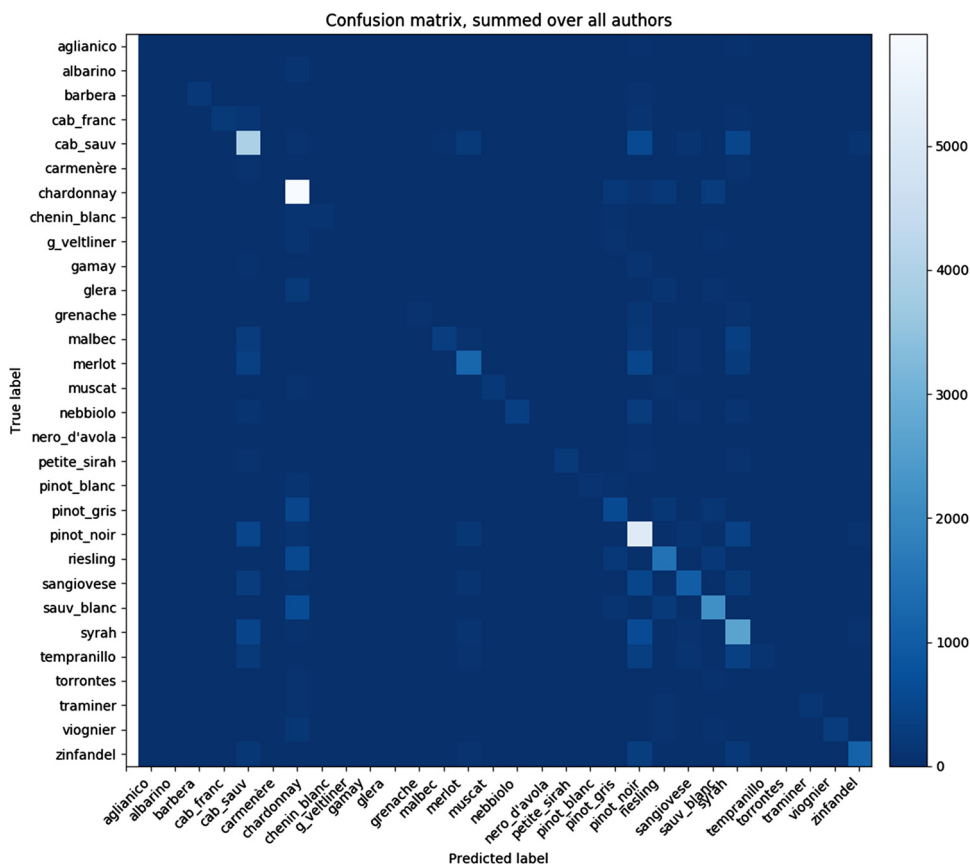


Figure 2. Confusion matrix for the grape variety classification task. Color shading indicates the relative number of individual classifications per cell with more classifications indicated by lighter cells.

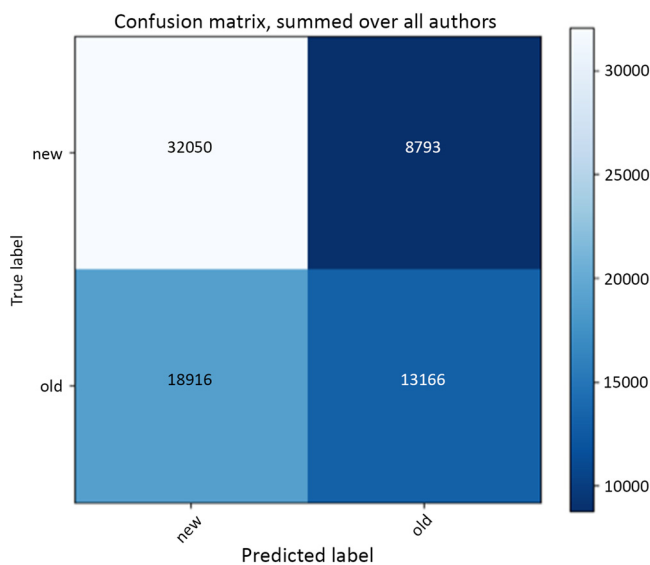


Figure 3. Confusion matrix for the old world–new world wine classification task. Color shading indicates the relative number of individual classifications per cell with more classifications indicated by lighter cells.

**Table 6.** Results per author for the new world versus old world wine classification task

| Author    | New world class   |         | Old world class   |         | Average |
|-----------|-------------------|---------|-------------------|---------|---------|
|           | Number of Reviews | F-score | Number of Reviews | F-score | F-score |
| Author 1  | 19266             | 84.2    | 90                | 1.9     | 72.7    |
| Author 2  | 65                | 1.3     | 11879             | 33.3    | 20.4    |
| Author 3  | 5428              | 68.7    | 4408              | 44.5    | 60.0    |
| Author 4  | 0.0               | 0.0     | 8622              | 80.4    | 67.2    |
| Author 5  | 5877              | 88.6    | 87                | 6.9     | 79.7    |
| Author 6  | 3075              | 71.5    | 2695              | 48.5    | 63.3    |
| Author 7  | 2306              | 76.8    | 1456              | 59.4    | 70.5    |
| Author 8  | 0.0               | 0.0     | 1770              | 90.3    | 82.3    |
| Author 9  | 1548              | 89.3    | 0.0               | 0.0     | 80.7    |
| Author 10 | 1190              | 85.0    | 87                | 23.0    | 74.9    |
| Author 11 | 844               | 71.8    | 301               | 42.3    | 62.1    |
| Author 12 | 794               | 71.9    | 260               | 43.8    | 62.5    |
| Author 13 | 450               | 62.9    | 427               | 61.6    | 62.3    |
| Total     | 40843             | 68.9    | 32082             | 49.4    | 61.5    |

In a first step, a list of candidate terms was generated from the corpus of wine reviews using part-of-speech pattern selection (i.e., nouns, adjectives, and verbs were included). Second, this list of terms was pruned by means of the Termhood weighting measure as implemented by Vintar (2010), such that the frequency of the candidate term was aligned with the frequency of that term in a background corpus: the Web 1T 5-gram v1 corpus. This corpus, made available by Google Inc., contains approximately one trillion word tokens from publicly accessible web pages (Brants and Franz 2006). The Termhood (T) term weighting measure of Vintar is computed as follows:

$$T(a) = \frac{F_a^2}{n} \sum_1^n \left( \log \frac{F_{n,D}}{N_D} - \log \frac{F_{n,R}}{N_R} \right) \quad (4)$$

in which  $F_a$  is the absolute frequency of the candidate term  $a$  in the (specialized) extraction corpus,  $F_{n,D}$  and  $F_{n,R}$  are the frequencies of each word in the extraction and in the general reference corpus, respectively, and  $N_D$  and  $N_R$  are the sizes of these two corpora expressed in the number of tokens.

The 1000 word  $n$ -grams or terms ranked highest by Termhood values for each author were concatenated into a single list of 13000 terms, and Termhood values were added for each author where possible, resulting in a 13000 term by 13 author matrix. Using 1000 words for each author gives greater opportunity for the lists of most frequent domain-specific terms to overlap, thus possibly inflating the rate of agreement. So, the same analysis was performed with only the first 100 terms ranked by Termhood values for each author. Most terms in both matrices were single words, but some bigrams also occurred (e.g., *green apple*, *dried fruit*).

The resulting matrices were used as input for PCA using R packages<sup>c</sup> FactoMineR (Lê *et al.* 2008) and factoextra (Kassambara and Mundt 2016). PCA is a technique that can be used to summarize and visualize (highly) multivariate data. This is done by maximizing the explained variance in the data from a number of variables and summarizing the data into components (Ringnér 2008). These components can subsequently be plotted in *n*-dimensional space using a visualization method, which here was the R package factoextra. For the current study, the different lists of unique terms per author with the Termhood values belonging to these terms are used as input for the PCA analysis. These lists may be different per author, or they may show overlap. The more inconsistent the descriptions of wines, the more different these Termhood lists would be per author, and the resulting PCA would produce a solution with many factors explaining the variance (i.e., potentially as many as the number of authors, i.e., 13). In contrast, if authors are consistent with their expert peers, the words in the Termhood ranked lists are expected to be very similar across authors, resulting in a solution with few or only a single factor.

### 3.2 Results

#### 3.2.1 Domain-specific wine vocabulary

Duplicate terms were removed from the concatenated list of 13,000 words, leaving 7853 unique terms. There was approximately 79.2% overlap in the terms used, with 5147 terms used frequently by at least two different authors.

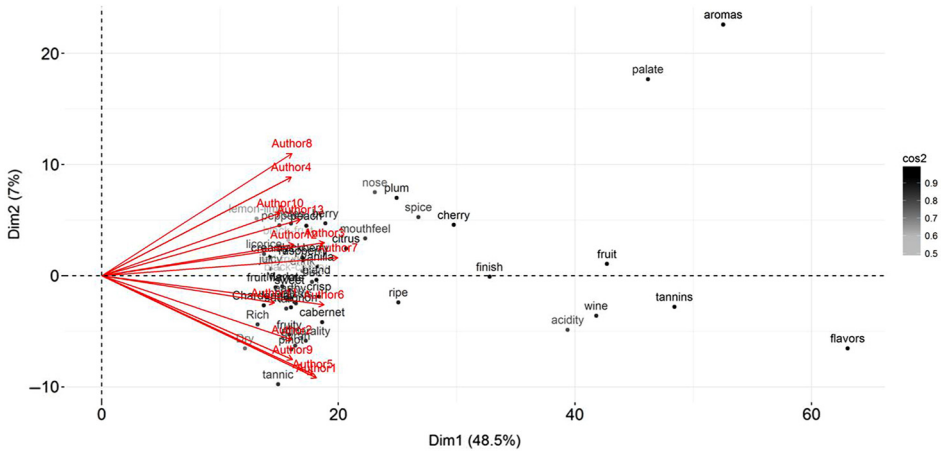
A scaled dual-factor PCA was performed over the 7853 unique terms by 13 author matrix. A scree plot, where factors are plotted according to the amount of variance they explain, was used to determine the amount of relevant factors. Scree plots normally give a distinct break between factors explaining a large part of the variance, and thus should be retained, and factors that do not explain a significant part of the variance (i.e., the scree, or rubbish; (Cattell 1966)). Inspection of the scree plot supported retaining a two-factor solution, although the eigenvalues suggested only the first factor was sufficient (eigenvalues: factor 1 = 6.51; factor 2 = 0.91; factor 3 = 0.80). To ease interpretation, the first two factors were retained. The first dimension explained 48.5% of the variance, and the second dimension 7.0%. All authors loaded positively on the first dimension (see Figure 4), suggesting consistency of term usage. According to the term loadings, the first dimension distinguished more general terms (e.g., *flavors*, *aroma*, *palate*), from more specific terms (e.g., *spice*, *vanilla*, *plum*, *lemon*). Authors seemed to be distinguished by the second dimension, with Author 8 and Author 5 being the most distinct from each other (see Figure 4). The second dimension also differentiated aroma terms from flavor terms; for example, terms like *plum* and *spice* loaded positively toward aromas, whereas words like *acidity* and *tannic* loaded negatively toward flavors.

To summarize, the solution was highly unidimensional, and all authors loaded positively on the first dimension. This suggests high consistency between authors in their language use. The authors nevertheless differed somewhat on the second dimension, suggesting some subtle differences in the use of aroma versus flavor terms.

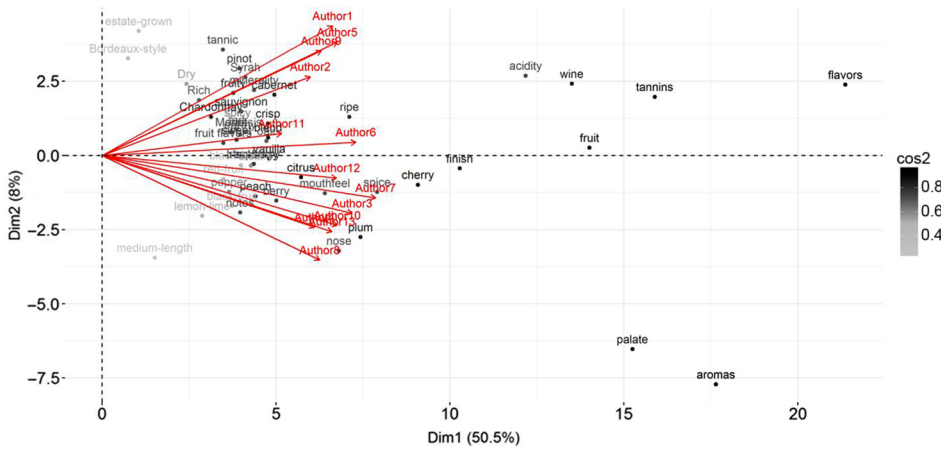
The same analysis was repeated with the first 100 terms ranked highest on Termhood for each author. There were 573 unique terms, with 96.4% of the terms used by at least two authors, and 146 terms used by all authors, that is, 146 terms were found to be used more frequently by every author compared to the reference corpus, as was indicated by the fact these terms had a positive Termhood value for these authors. One could conclude that there are 146 “wine terms,” that is, they are used distinctly (compared to the use of those terms in Standard English) and conventionally (used across wine writers).

The result of this second PCA was similar to the first. The eigenvalues (eigenvalues: factor 1 = 6.57; factor 2 = 1.05; factor 3 = 0.85) and scree plot suggested a two-factorial solution. The

<sup>c</sup><https://www.r-project.org>



**Figure 4.** Biplot of PCA analysis conducted on the Termhood weighted wordlists ( $n = 1000$ ) for each author. Terms are shown as cases, grey-scaled by their relative contribution toward the solution (cos2 weighed; (Abdi and Williams 2010)), and authors are shown in red. Red vectors indicate the correlation between both dimensions for each author. To ease interpretation, only the 50 most influential terms in the solution are plotted in this graph.

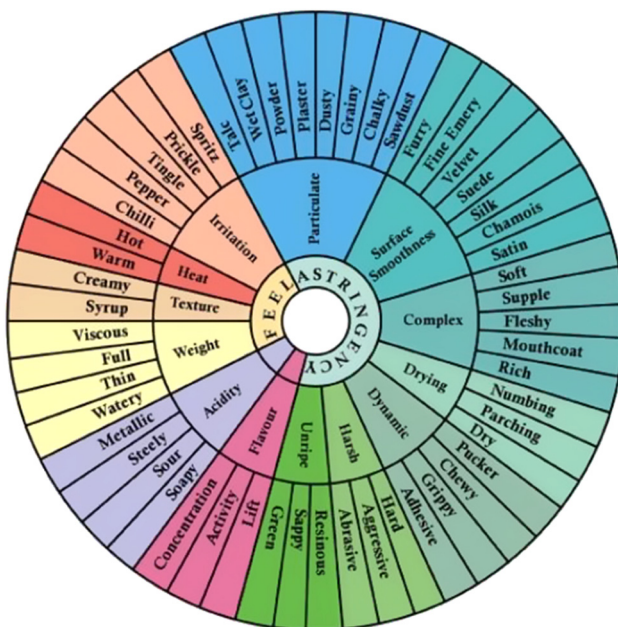


**Figure 5.** Biplot of PCA analysis conducted on the Termhood weighed wordlists ( $n = 100$ ) for each author. Terms are shown as cases, colored by their relative contribution toward the solution (cos2 weighed; (Abdi and Williams 2010)), and authors are shown in red. Red vectors indicate the relative correlation both dimensions for each author. To ease interpretation, only the 50 most influential terms in the solution are plotted in this graph.

first dimension explained 50.5% of the data, and the second dimension 8.0% (Figure 5). Authors loaded positively and with comparable influence on the first dimension (shown by the red vectors in Figure 5). The first dimension ranged roughly from specific words (*peach, crisp, vanilla, pinot noir*) to more general words (*flavors, fruit, palate, aromas*). The second dimension was reversed with respect to the first PCA analysis, that is, ranged from flavors to aromas, but as the scale of PCA factors is arbitrarily determined, it is comparable to the first analysis. The authors showed some dispersion on this second dimension, with positive loadings for Author 1 and negative loadings for Author 8, on the extremes.

To summarize, the PCA analyses further confirmed that authors are generally consistent with each other in their descriptions. The first dimension of the PCA solution revealed consensus between authors, and ranged from specific to general terms. The second dimension showed some





**Figure 6.** Mouthfeel terminology wheel showing a hierarchical representation of terms used to describe the mouthfeel of red wine. Adapted with permission from Gawel *et al.* (2000).

dispersion between authors. However, the variance explained by this dimension was small. Terms used to indicate flavors, including aspects such as taste, or grape type loaded highly on one end of the second dimension, while source terms referring to aromas such as plum loaded on the other end of this dimension. This suggests that while authors were remarkably consistent overall, authors differed somewhat in their strategy to describe wines by taking either a more flavor-driven approach, for example, Author 1 and Author 5, or a more aroma-driven approach exemplified by Author 8.

### 3.2.2 Comparison of wine vocabulary

Previously, scholars have compiled lists of wine vocabulary. Notably, Lehrer (2009) describes three wine wheels, that is, the aroma wheel (Noble *et al.* 1984, 1987), the sparkling wine wheel (Noble and Howe 1990), and the mouthfeel terminology wheel (see Figure 6; Gawel *et al.* (2000)). As introduced before, a wine wheel is a list of terms that can be used to describe a wine, organized by specificity: the most general terms are listed on the middle tier, and more specific words are listed on outer tiers (see Figure 6, for example). We compiled words from these three classic wine wheels which resulted in a single list of 244 unique terms. In addition, two other vocabulary lists: (i) Robert Parkers' glossary of 117 wine terms (Parker 2017), and (ii) the 61 references used in the Le Nez du Vin wine aroma kit (Lenoir 2011) were collated. The Le Nez du Vin Masterkit contains 54 labeled smells. These were supplemented with the 12 reference terms from the New Oak kit. After removal of duplicate terms that occurred in both kits, 61 terms remained. These existing lists of wine vocabulary were compared to the domain-specific vocabulary that was found in the current corpus of wine reviews.

Before comparing the vocabularies in the various wheels and lists, we first built our own wine wheel—the *Text-Based Wine Wheel*—which visualized the terms we extracted from our wine corpus in a completely bottom-up manner (see Figure 7). To construct our wheel, the 146 unique terms extracted using TExSIS were organized on a wine wheel using the *XLStat Sensory Wheel* function. After minimal preprocessing (e.g., *spice* and *spicy* were combined into one entry), the automatically extracted terms of the outer ring were manually classified into 3 overarching categories and 12 subcategories that are depicted on the inner rings: aromas (fruit; spices; food;

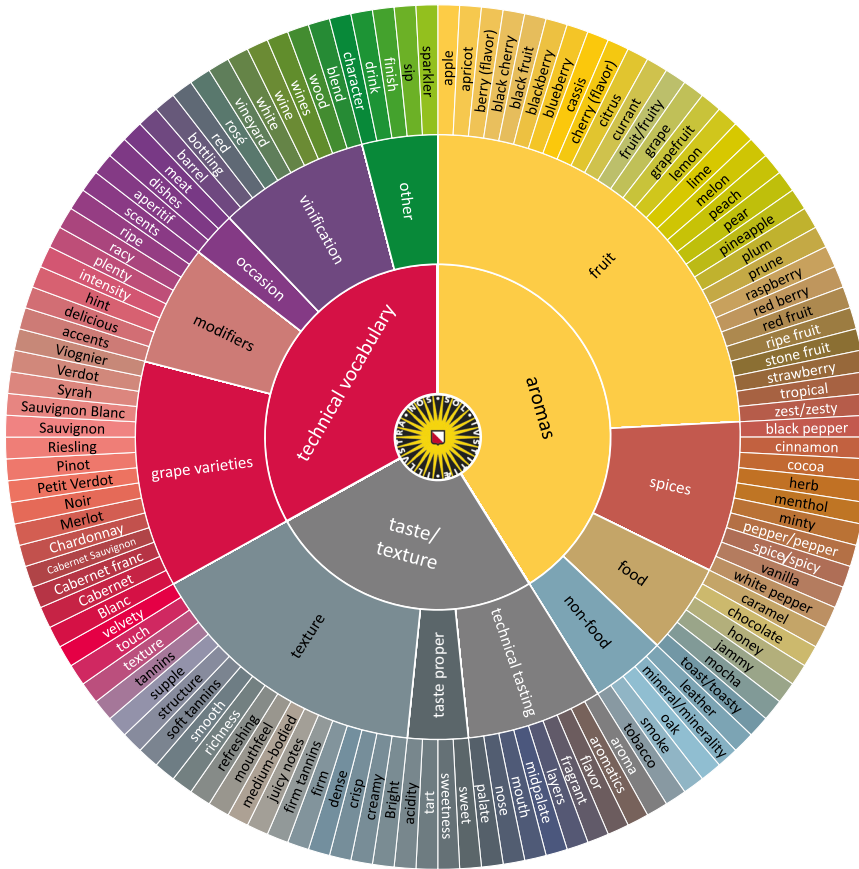


Figure 7. The *Text-Based Wine Wheel*, based on the terms automatically extracted from our corpus of wine expert reviews (outer ring), and grouped into categories (inner rings).

non-food), taste/texture (technical tasting; taste proper; texture), and technical vocabulary (grape varieties; modifiers; occasion; vinification; other).

We then compared the wine vocabularies from the various sources. First, the terms in each list (the 244 terms from classic wine wheels, 117 terms from Robert Parker's glossary, and 61 references from the Le Nez du Vin Masterkit, in addition to the list of 146 terms uncovered from the present corpus: our Termhood list) were further processed. Spelling variants were standardized. Some lemmas had multiple entries, for example, the singular *fruit* and plural *fruits*, and were collapsed. Adverbial phrasings such as *fruity* possibly apply to more distinct smells than *fruit*, so these were kept separate, as was *cherry flavors*, which possibly covers more flavors than *cherry* alone. Also, *drying* and *dry* were kept as unique entries. Next, the vocabulary from the classic wine wheels was examined qualitatively to determine the amount of overlap (Table 7). Out of the 244 terms that occurred in at least one of the three classic wine wheels, 34 also appeared in our list (i.e., 13.9% overlap); 13 terms occurred in both Parker's glossary and our list (i.e., 11.1%), while 21 overlapped in the Le Nez du Vin reference list and ours (almost 30% of the 61 terms on the Le Nez du Vin list). In total, 45 terms occurred in all lists. Not only does this suggests some overlap, but also that there are many words listed in wine vocabulary lists that are not frequently used in actual wine descriptions (at least in our data). One possibility is that the words not attested in our corpus of wine reviews denote very specific aromas and flavors not commonly found across a range of wines. Nevertheless, both novices and experts may also benefit from a list of the more common vocabulary from wine reviews. The new *Text-Based Wine Wheel* provides just such a tool.

**Table 7.** Words occurring both in the Termhood highest ranked list and in the established wine vocabulary list

| Terms overlapping in the Termhood list and classic wine wheels ( <i>n</i> = 34)   | Terms overlapping in the Termhood list and Parker's glossary ( <i>n</i> = 13)                        | Terms overlapping in the Termhood list and the Le Nez Du Vin list ( <i>n</i> = 21)   |
|---|--|--|
| Acidity; apple; apricot; berry; black pepper; blackberry; caramel; cherry; chocolate; cinnamon; citrus; cocoa; creamy; fruity; grapefruit; honey; lemon; lime; melon; menthol; oak; peach; pear; pepper; pineapple; prune; raspberry; spicy; strawberry; supple; texture; tobacco; tropical; fruit; vanilla | Acidity; aroma; intensity; jammy; nose; peppery; ripe; spicy; supple; tart; toasty; tobacco; velvety | Apple; apricot; blackberry; caramel; cherry; cinnamon; grapefruit; honey; leather; lemon; melon; oak; peach; pear; pepper; pineapple; prune; raspberry; strawberry; toast; vanilla |

Of further interest are the unique terms in the Termhood list. These 89 terms are used often in online wine reviews; in fact, they were used by all 13 authors, and all with higher frequency than there are likely to occur in everyday English; yet, these terms are not included in reference word lists such as the Noble wine wheel and Parker's glossary. These were aroma terms such as *black cherry, blueberry, cassis, cherries, cocoa, fruit, lime, mocha, red berry, red fruit, ripe fruit, smoke, spice, stone fruit, tannins, wood, zest*. Some of these words were adjectives, that is, *bright, creamy, crisp, delicious, dense, firm, juicy, minty, racy, smooth, zesty*; while other terms picked out intensity or complexity, such as *accents, layers, hint, notes, plenty, richness, scents*. Other terms indicated the location or modality in which the flavor was perceived, that is, *finish, midpalate, mouth, mouthfeel, palate, sweet, structure, touch*. In addition, a number of terms can be considered technical language about grape types, vinification methods, and comments on how to enjoy the wine best, for example, *blend, cabernet sauvignon, merlot, riesling, viognier, vineyard, minerality, aperitif, dishes* (see Supplementary materials S1 for full lists).

**4. Discussion**

Controversy surrounds expert descriptions of wine. On the one hand, tasting notes are criticized, and described as uninformative (Shesgreen 2003; Quandt 2007) and highly idiosyncratic (Lawless 1984). The current results contradict these proposals. Wine reviews were found to consistently distinguish global properties of wine, such as color and grape variety. The fact that reviewers—despite their individual vocabulary preferences—distinguish categories of wines consistently is impressive. Wine experts are able to write in their individual styles while at the same time giving consistent descriptions of wine.

Solomon (1997) proposed that when novices become wine experts, they undergo a conceptual shift, that is, their knowledge structures become more refined, and the conceptual categories become more specific (Carey 2000). He further hypothesized that wine expert knowledge is organized by grape type. Later studies have shown wine experts indeed consistently sort wines by grape type, while novices use other (more haphazard) strategies (Solomon 1997; Ballester *et al.* 2008; Urdapilleta *et al.* 2011). The current study shows expert language also distinguishes grape varieties distinctively, further corroborating the hypothesis that wine knowledge is structured by grape variety.

Wine is highly multidimensional. In addition to grape variety, the color of wine affects how experts describe wines (Morrot, Brochet and Dubourdieu 2001; Parr, White and Heatherbell 2003) and color can influence how sweet a wine is perceived to be (Pangborn, Berg and Hansen 1963). When experts do not taste wines blind, their perception and descriptions are influenced by what they see (Auvray and Spence 2008; Smith 2012; Spence 2015*b*). In the current study, the color of wine was also reflected in descriptions from experts, further underlining the importance of color, and vision in general, in flavor perception (Auvray and Spence 2008; Christensen 1983).

We hypothesized wine experts would also vary their descriptions of wines by origin. A recent study suggested terroir, that is, the place where wine is made, has a bigger influence on the smell of a wine than grape type (Foroni *et al.* 2017). In the current study, we investigated a coarse-grained distinction by examining whether reviews distinguished wine made in the old or new world. This distinction is often made by wine experts (Remaud and Couderc 2006), but has received criticism too (Remaud and Couderc 2006; Banks and Overton 2010). In line with this criticism, the classification task did not yield reliable results, which may indicate the old versus new world distinction is not consistently reflected in wine experts' descriptions, and further suggests experts might not think about wines along this dimension. Nevertheless, a more fine-grained distinction, on the level of country of origin, or even at the level of the specific wine region, may be important and is worthy of further examination.

Wine writers write not only for other experts but also for less knowledgeable consumers. There is little relationship between price and quality for wines (Cardebat and Livat 2016; Oczkowski and Doucouliagos 2015), so reviews can provide important guides for the less experienced. In order to become a wine expert, students have to practice naming the aromas and flavors that can be encountered in a wine. A structured list of words, such as the Wine wheel (Noble *et al.* 1984), can be a useful tool to help budding wine enthusiasts to develop their ability to describe wines. In fact, wine wheels such as those from Noble *et al.* (1987), as well as the vocabulary used by Robert Parker, have changed the way wines are described dramatically over the last 50 years (James 2018). At the same time, there is criticism of the use of these vocabulary lists. Most of the terms are not exclusive to the domain of olfaction and are essentially metaphors from other domains. Novices, lacking appropriate background, may struggle to understand what is meant by a descriptor from these lists (Lawless 1984; James 2018). Nevertheless, conventionalized vocabulary may help professionals to standardize their descriptions, improve communicative efficacy, and may also aid in tailoring descriptions for less knowledgeable consumers (Gawel *et al.* 2000). In line with the suggestion of Gawel *et al.* (2000), that wine vocabularies should be frequently revisited and updated, we show that many words used by wine experts were not present on these lists. This suggests there is still room for improvement.

Nevertheless, the current list of words has caveats too. One striking observation is that most words are positive, and none of the words can be considered negative. This contrasts the wine vocabulary of Robert Parker, for example, where many words such as *dumb*, *closed*, and *off* can be used to describe wines more negatively. The data that were used here to obtain the vocabulary could explain why most of the found terms were positive: negative reviews are rarely published on the source website, and all wines in the database scored above 75 out of 100 points. Future follow-up investigations may consider texts obtained from naming experiments done with experts, with a broad qualitative range of wines, which includes wines with wine faults (Lawless 1984; Solomon 1990; Melcher and Schooler 1996; Solomon 1997; Croijmans and Majid 2016). Although this approach would result in a significantly smaller data set, it may nevertheless be a valuable supplement to the wine vocabulary found here, and in other existing wine vocabularies.

To conclude, in other expertise domains, such as dog breeding and bird watching (Tanaka and Taylor 1991), computer maintenance (Humphrey and Underwood 2011), and the visual arts (Cialone *et al.* 2018), expertise has been shown to affect the use of visual and spatial language. The current results show that even in a domain that is incredibly difficult to talk about for the general population—that is, olfaction—expertise can shape language use. Wines were described using domain-specific language in a consistent and distinct manner. This shows wine experts can overcome the limitations of their language (Levinson and Majid 2014) and convey experiences of smell and flavor with verve.

**Acknowledgements.** This work was funded by The Netherlands Organization for Scientific Research: NWO VICI grant “Human olfaction at the intersection of language, culture and biology” to A. Majid [grant number 277-70-011]. Thanks to Chris van der Lee for processing the corpus data, and Laura Speed and Artin Arshamian for comments on an earlier draft of the manuscript. We would like to thank three anonymous reviewers for their thoughtful comments and suggestions on two earlier drafts of this manuscript.



## References

- Abdi H. and Williams L.J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(4), 433–459.
- Auvray M. and Spence C. (2008). The multisensory perception of flavor. *Consciousness and Cognition* 17(3), 1016–1031.
- Ballester J., Patris B., Symoneaux R. and Valentin D. (2008). Conceptual vs perceptual wine spaces: Does expertise matter? *Food Quality and Preference* 19, 267–276.
- Banks G. and Overton J. (2010). Old world, new world, third world? Reconceptualising the worlds of wine. *Journal of Wine Research* 21(1), 57–75.
- Biederman I. and Shiffrar M.M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13(4), 640.
- Boesveldt S. and de Graaf K. (2017). The differential role of smell and taste for eating behavior. *Perception* 46(3–4), 307–319.
- Brants T. and Franz A. (2006). Web 1t 5-gram version 1. *Linguistic Data Consortium, Philadelphia*.
- Brochet F. and Dubourdieu D. (2001). Wine descriptive language supports cognitive specificity of chemical senses. *Brain and Language* 77(2), 187–196.
- Brodsky W., Henik A., Rubinstein B. and Zorman M. (2003). Auditory imagery from musical notation in expert musicians. *Perception and Psychophysics* 65(4), 602–612.
- Burenhult N. and Majid A. (2011). Olfaction in Asian ideology and language. *The Senses and Society* 6(1), 19–29.
- Caballero, R. and Suárez-Toste E. (2010). A genre approach to imagery in winespeak: Issues and prospects. *Researching and Applying Metaphor in the Real World* 26, 265–288.
- Cain W.S. (1979). To know with the nose: Keys to odor identification. *Science* 203(4379), 467–470.
- Cain W.S., de Wijk R., Lulejian C., Schiet F. and See L. (1998). Odor identification: Perceptual and semantic dimensions. *Chemical Senses* 23(3), 309–326.
- Caley M.J., O’Leary R., Fisher R., Low-Choy S., Johnson S. and Mengersen K. (2014). What is an expert? A systems perspective on expertise. *Ecology and Evolution* 4(3), 231–242.
- Cardebat J. and Livat F. (2016). Wine experts’ rating: A matter of taste? *International Journal of Wine Business Research*, 28(1), 43–58.
- Carey S. (2000). The origin of concepts. *Journal of Cognition and Development* 1(1), 37–41.
- Cattell R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research* 1(2), 245–276.
- Chang C. and Lin C. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 27:1–27:27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Christensen C.M. (1983). Effects of color on aroma, flavor and texture judgments of foods. *Journal of Food Science* 48(3), 787–790.
- Cialone C., Tenbrink T. and Spiers H.J. (2018). Sculptors, architects, and painters conceive of depicted spaces differently. *Cognitive Science* 42(2), 524–53.
- Croijmans I. and Majid A. (2016). Not all flavor expertise is equal: The language of wine and coffee experts. *PLoS ONE* 11(6), e0155845.
- de Groot A.D. (1978). *Thought and Choice in Chess*, 2nd Edn. The Hague, The Netherlands: Mouton Publishers.
- De Groot A.D., Gobet F. and Jongman R.W. (1996). *Perception and Memory in Chess: Studies in the Heuristics of the Professional Eye*. Assen, The Netherlands: Van Gorcum & Co.
- De Groot A.D. (1946). *Het denken van den schaker: een experimenteel-psychologische studie*. Amsterdam, The Netherlands: Noord-Hollandsche Uitgevers Maatschappij.
- De Valk J.M., Wnuk E., Huisman J.L.A. and Majid A. (2017). Odor–color associations differ with verbal descriptors for odors: A comparison of three linguistically diverse groups. *Psychonomic Bulletin and Review* 24(4), 1171–1179.
- Engen T. (1987). Remembering odors and their names. *American Scientist* 75(5), 497–503.
- Ericsson K.A., Hoffman R.R., Kozbelt A. & Williams A.M. (eds). (2018). *The Cambridge Handbook of Expertise and Expert Performance*, 2nd Edn. Cambridge, United Kingdom: Cambridge University Press.
- Ericsson, K.A., Prietula, M.J. and Cokely, E.T. (2007). The making of an expert. *Harvard Business Review* 85(7/8), 1–8.
- Faroni F., Vignando M., Aiello M., Parma V., Paoletti M.G., Squartini A. and Rumiati R.I. (2017). The smell of terroir! Olfactory discrimination between wines of different grape variety and different terroir. *Food Quality and Preference* 58, 18–23.
- Fujii N., Abia D., Kudo N., Hihara S., Okanoya K. and Iriki A. (2007). Prefrontal activity during koh-do incense discrimination. *Neuroscience Research* 59(3), 257–264.
- Gawel R. (1997). The use of language by trained and untrained experienced wine tasters. *Journal of Sensory Studies* 12(4), 267–84.
- Gawel R., Oberholster A. and Francis I.L. (2000). A ‘Mouth-feel Wheel’: Terminology for communicating the mouth-feel characteristics of red wine. *Australian Journal of Grape and Wine Research* 6(3), 203–207.
- Gluck M. (2003). Chapter 11: Wine language. Useful idiom or idiot-speak. In Aitchison J. and Lewis D. (eds), *New Media Language*. London, United Kingdom: Routledge, pp. 107–115.

- Halpern A.R. and Bower G.H. (1982). Musical expertise and melodic structure in memory for musical notation. *The American Journal of Psychology* 95(1), 31–50.
- Hendrickx I., Lefever E., Croijmans I., Majid A. and van den Bosch A. (2016). Very quaffable and great fun: Applying NLP to wine reviews. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, pp. 306–312.
- Herdenstam A.P.F., Hammarén M., Ahlström R. and Wiktorsson P. (2009). The professional language of wine: Perception, training and dialogue. *Journal of Wine Research* 20(1), 53–84.
- Humphrey K. and Underwood G. (2011). See what I'm saying? expertise and verbalisation in perception and imagery of complex scenes. *Cognitive Computation* 3(1), 64–78.
- James A. (2018). How Robert Parkers 90+ and Ann Noble's aroma wheel changed the discourse of wine tasting notes. *ILCAE. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 31.
- Joachims T. (2002). *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Juola P. (2008). Authorship attribution. *Foundations and Trends in Information Retrieval* 1(3), 233–334.
- Kageura K. and Umino B. (1996). Methods of automatic term recognition: A review. *Terminology* 3(2), 259–89.
- Kassambara A. and Mundt F. (2016). Factoextra: Extract and visualize the results of multivariate data analyses. *R Package Version 1.0.3*. <https://rdr.io/cran/factoextra/>.
- Kestemont M., Luyckx K., Daelemans W. and Crombez T. (2012a). Cross-genre authorship verification using unmasking. *English Studies* 93(3), 340–356.
- Kestemont M., Daelemans W. and Sandra D. (2012b). Robust rhymes? The stability of authorial style in medieval narratives. *Journal of Quantitative Linguistics* 19(1), 54–76.
- Lawless H.T. (1984). Flavor description of white wine by “expert” and nonexpert wine consumers. *Journal of Food Science* 49(1), 120–123.
- Lê S., Josse J. and Husson F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software* 25(1), 1–18.
- Lehrer A. (2009). *Wine and Conversation*, 2nd Edn. New York: Oxford University Press.
- Lenoir J. (2011). *Le nez du vin*. Editions Jean Lenoir.
- Levinson S.C. and Majid A. (2014). Differential ineffability and the senses. *Mind and Language* 29(4), 407–427.
- Levitin D.J. and Rogers S.E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences* 9(1), 26–33.
- Lorig T.S. (1999). On the similarity of odor and language perception. *Neuroscience and Biobehavioral Reviews* 23(3), 391–398.
- Macken L., Lefever E. and Hoste V. (2013). TExSIS: Bilingual terminology extraction from parallel corpora using chunk-based alignment. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 19(1), 1–30.
- Majid A. (2015). Cultural factors shape olfactory language. *Trends in Cognitive Sciences* 19(11), 629–630.
- Majid A. and Burenhult N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition* 130(2), 266–270.
- Majid A. and Kruspe N. (2018). Hunter-gatherer olfaction is special. *Current Biology* 28(3), 409–13.
- Majid A., Roberts S.G., Cilissen L., Emmorey K., Nicodemus B. and Levinson S.C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences of the United States of America* 115(45), 11369–11376.
- Manning C.D., Surdeanu M., Bauer J., Finkel J., Bethard S.J. and McClosky D. (2014). The Stanford CoreNLP Natural Language Processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60. Baltimore, Maryland USA, June 23–24, 2014.
- Melcher J.M. and Schooler J.W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language* 35(2), 231–245.
- Mitchell H.F. and MacDonald R.A.R. (2011). Remembering, recognizing and describing singers' sound identities. *Journal of New Music Research* 40(1), 75–80.
- Morrot G., Brochet F. and Dubourdieu D. (2001). The color of odors. *Brain and Language* 79(2), 309–320.
- Noble A.C., Arnold R.A., Buechsenstein J., Leach E.J., Schmidt J.O. and Stern P.M. (1987). Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture* 38(2), 143–146.
- Noble A.C., Arnold R.A., Masuda B.M., Pecore S.D., Schmidt J.O. and Stern P.M. (1984). Progress towards a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture* 35(2), 107–109.
- Noble A.C., and Howe P. (1990). *The Sparkling Wine Aroma Wheel*. Davis, CA. as cited in Lehrer, 2009.
- Oczkowski E. and Doucouliagos H. (2015). Wine prices and quality ratings: A meta-regression analysis. *American Journal of Agricultural Economics* 97(1), 103–121.
- Olofsson J.K. and Gottfried J.A. (2015). The muted sense: Neurocognitive limitations of olfactory language. *Trends in Cognitive Sciences* 19(6), 314–321.
- O'Meara C. and Majid A. (2016). How changing lifestyles impact Seri smellscape and smell language. *Anthropological Linguistics* 58(2), 107–131.



- Pangborn R.M., Berg H.W. and Hansen B. (1963). The influence of color on discrimination of sweetness in dry table-wine. *The American Journal of Psychology* 76(3), 492–495.
- Paradis C. and Eeg-Olofsson M. (2013). Describing sensory experience: The genre of wine reviews. *Metaphor and Symbol* 28(1), 22–40.
- Parker R. (2017). Glossary Terms. <https://www.robertparker.com/resources/glossary-terms> (accessed 15 June 2017).
- Parr W.V., Mouret M., Blackmore S., Pelquest-Hunt T. and Urdapilleta I. (2011). Representation of complexity in wine: Influence of expertise. *Food Quality and Preference* 22(7), 647–660.
- Parr W.V., White G.K. and Heatherbell D.A. (2003). The nose knows: Influence of colour on perception of wine aroma. *Journal of Wine Research* 14(2–3), 79–101.
- Pluijms J.P., Cañal-Bruland R., Bergman Tiest W.M., Mulder F.A. and Savelsbergh G.J.P. (2015). Expertise effects in cutaneous wind perception. *Attention, Perception, and Psychophysics* 77(6), 2121–2133.
- Quandt R.E. (2007). On wine bullshit: Some new software? *Journal of Wine Economics* 2(02), 129–135.
- Remaud H. and Couderc J. (2006). Wine business practices: A new versus old wine world perspective. *Agribusiness: An International Journal* 22(3), 405–416.
- Ringnér M. (2008). What is principal component analysis? *Nature Biotechnology* 26(3), 303–304.
- Rivlin R. and Gravelle K. (1984). *Deciphering the Senses: The Expanding World of Human Perception*. New York: Simon and Schuster.
- Royet J., Plailly J., Saïve A., Veyrac A. and Delon-Martin C. (2013). The impact of expertise in olfaction. *Frontiers in Psychology* 4, 928–939.
- San Roque L., Kendrick K.H., Norcliffe E., Brown P., Defina R., Dingemanse M., Dirksmeyer T., Enfield N.J., Floyd S., Hammond J., Rossi G., Tufvesson S., Van Putten S. and Majid A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics* 26(1), 31–60.
- Sauvageot F., Urdapilleta I. and Peyron D. (2006). Within and between variations of texts elicited from nine wine experts. *Food Quality and Preference*, 17(6), 429–444.
- Shepherd G.M. (2006). Smell images and the flavour system in the human brain. *Nature* 444(7117), 316–321.
- Shesgreen S. (2003). Wet dogs and gushing oranges: Winespeak for a new millennium. *The Chronicle of Higher Education* 49(7), 572–575.
- Silverstein M. (2006). Old wine, new ethnographic lexicography. *Annual Review of Anthropology* 35(1), 481–496.
- Smith B. (2012). Perspective: complexities of flavour. *Nature* 486(7403), S6–S6.
- Solomon G.E.A. (1990). Psychology of novice and expert wine talk. *The American Journal of Psychology* 103(4), 495–517.
- Solomon G.E.A. (1997). Conceptual change and wine expertise. *The Journal of the Learning Sciences* 6(1), 41–60.
- Sowden P.T., Davies I.R.L. and Roling P. (2000). Perceptual learning of the detection of features in X-ray images: A functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance* 26(1), 379–390.
- Spence C. (2015a). Just how much of what we taste derives from the sense of smell? *Flavour* 4(1), 30.
- Spence C. (2015b). Multisensory flavor perception. *Cell* 161(1), 24–35.
- Sperber D. (1975). *Rethinking symbolism*. Cambridge, England: Cambridge University Press (Original work published in 1974).
- Suárez Toste E. (2007). Metaphor inside the wine cellar: On the ubiquity of personification schemas in winespeak. *Metaphorik.de* 12(1), 53–64.
- Tanaka J.W. and Taylor M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 23(3), 457–482.
- Urdapilleta I., Parr W., Dacremont C. and Green J. (2011). Semantic and perceptive organisation of sauvignon blanc wine characteristics: Influence of expertise. *Food Quality and Preference* 22(1), 119–128.
- Van Rijsbergen C.J. (1979). *Information Retrieval*. London: Butterworth.
- Vintar S. (2010). Bilingual term recognition revisited. The bag-of-equivalents term alignment approach. *Terminology* 16(2), 141–158.
- Weinstein B.D. (1993). What is an expert? *Theoretical Medicine* 14(1), 57–73.
- Wnuk E. and Majid A. (2014). Revisiting the limits of language: The odor lexicon of Maniq. *Cognition* 131(1), 125–138.
- Yeshurun Y. and Sobel N. (2010). An odor is not worth a thousand words: From multidimensional odors to unidimensional odor objects. *Annual Review of Psychology* 61(1), 219–241.
- Zheng R., Li J., Chen H. and Huang Z. (2006). A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology* 57(3), 378–793.

**Cite this article:** Croijmans I, Hendrickx I, Lefever E, Majid A and Van Den Bosch A (2020). Uncovering the language of wine experts. *Natural Language Engineering* 26, 511–530. <https://doi.org/10.1017/S1351324919000500>