White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/137244/

Version: Published Version

Proceedings Paper:
Lefever, Els, Hendrickx, Iris, Croijmans, Ilja et al. (2018) Discovering the Language of Wine Reviews: A Text Mining Account. In: Calzolari, Nicoletta, Choukri, Khalid, Cieri, Christopher, Declerck, Thierry, Goggi, Sara, Hasida, Koiti, Isahara, Hitoshi, Maegaard, Bente, Mariani, Joseph, Mazo, Hélène, Moreno, Asuncion, Odijk, Jan, Piperidis, Stelio and Tokunaga, Takenobu, (eds.) Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). LREC, Miyazaki,Japan, pp. 3297-3302.

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 including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Discovering the Language of Wine Reviews: A Text Mining Account

Els Lefever[1], Iris Hendrickx[2], Ilja Croijmans[2], Antal van den Bosch[2,4], Asifa Majid[2,3]
[1]Language and Translation Technology Team, Ghent University, Belgium
[2]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
[3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[4]Meertens Institute, Amsterdam, The Netherlands
els.lefever@ugent.be, {i.hendrickx,i.croijmans,a.vandenbosch,asifa.majid}@let.ru.nl

Abstract

It is widely held that smells and flavors are impossible to put into words. In this paper we test this claim by seeking predictive patterns in wine reviews, which ostensibly aim to provide guides to perceptual content. Wine reviews have previously been critiqued as random and meaningless. We collected an English corpus of wine reviews with their structured metadata, and applied machine learning techniques to automatically predict the wine's color, grape variety, and country of origin. To train the three supervised classifiers, three different information sources were incorporated: lexical bag-of-words features, domain-specific terminology features, and semantic word embedding features. In addition, using regression analysis we investigated basic review properties, i.e., review length, average word length, and their relationship to the scalar values of price and review score. Our results show that wine experts do share a common vocabulary to describe wines and they use this in a consistent way, which makes it possible to automatically predict wine characteristics based on the review text alone. This means that odors and flavors may be more expressible in language than typically acknowledged.

Keywords: wine reviews, wine vocabulary, classification, supervised learning, terminology extraction

1. Introduction

Few food categories are described as often as wine: a vast number of wine reviews appear in magazines, books, blogs, supermarkets, newspapers, and numerous other venues. These reviews contain descriptors, which chronicle the appearance, aroma (smell), flavor, and textural attributes of wines in loving detail. This is puzzling, as previously many scholars have claimed that smells and flavors are difficult, if not impossible to put into words (Sperber, 1975; Rouby et al., 2002; Levinson and Majid, 2014).

There is an ongoing debate as to whether wine is actually described in an informative manner in wine reviews. On the one hand, studies suggest wine experts use language in a consistent manner (Croijmans and Majid, 2016). Similarly, expert descriptions are more often correctly matched to a wine than descriptions written by novices (Solomon, 1990), with the suggestion that wine experts are more likely to use specific terminology which is more informative. On the other hand, studies suggest trained wine experts more often use vague and abstract terms (e.g., *complex, attractive*) when describing wines (Gawel, 1997). Similarly, metaphorical language is often encountered in wine descriptions (Suárez Toste, 2007; Caballero, 2007), suggesting wine experts employ vague and overly literary prose. Others have more directly critiqued wine reviews as being uninformative and mere "purple prose" (Quandt, 2007) (p.130).

Here, we study how wine experts express properties of wines in their reviews. Wine reviews convey both sensory descriptions of wines, as well as objective properties such as color, grape type, country of origin, and price; and reviews also convey an overall rating as to the quality of the wine. Previously, we showed that experts are indeed consistent in their descriptions and we were able to train a classifier on review texts to predict objective wine properties (color, country, grape, price) (Hendrickx et al., 2016). In the present investigation we expand this work in two directions.

First, we examined the usefulness of domain-specific terminology as feature representations for classification tasks. To investigate the terminological consistency in wine reviews, we set up a machine learning experiment to automatically predict the color, grape type, and country of origin of wines based on the information contained in review texts. To this end, we used a corpus of online wine reviews and their structured metadata and extracted three types of information from the review text: a set of lexical bag-of-words features, a set of domain-specific terminological features, and a set of semantic word embedding cluster features.

Second, we investigated a non-textual subjective property assigned to the wine, namely the rating that was given by the expert who wrote the review. We wished to establish whether there was a correlation between prices and ratings, and whether the rating also influenced the review text. Previous work suggests that wine experts (contrary to laypersons) prefer more expensive wines over cheaper ones (Goldstein et al., 2008). Furthermore, more expensive wines are described with longer reviews, measured in the total number of characters in the whole review (Ramirez, 2010). Another way in which price and rating may be reflected in the review may be in the average word length of the words used: More expensive wines may be described using "more expensive" words, i.e., longer words, for example. We performed a regression analysis to replicate the finding by Ramirez (2010) and to further explore the relationship between subjective ratings and price, and the length of wine reviews and average word length in the reviews.

The remainder of this paper is organized as follows: in Section 2 we give an overview of related research, while in Section 3 we provide details about our corpus of wine expert reviews. Section 4 describes the experimental setup and results of our classification experiments to predict wine color,

type, and country of origin. Section 5 presents the regression analysis where we focus on the aspects of price and rating. Section 6 summarizes our main findings.

2. Related research

A number of studies have covered related territory. For this abstract we briefly review some of the most pertinent studies.

Brochet and Dubourdieu (2001) carried out a lexical analysis of four corpora of wine tasting comments by performing χ^2 calculations for all word co-occurrences in the text. The resulting lexical fields were not organized along sensory dimensions only, but contained a mix of visual (*yellow*), olfactory (*apricot*), taste (*sweet*) and hedonistic (*good*) terms, among others, which appears to contradict reports of professional tasters who say they taste wine in an "analytic" manner. Moreover, the fact that word groups combine visual, olfactory and taste descriptors, support the idea that wine language is organized around wine "prototypes". Indeed, Solomon (1997) showed that features identified by experts significantly co-vary with grape types such that wines of the same grape are described more similarly by experts, suggesting these prototypes may be real.

Wine reviews also often feature ratings of wines (i.e., a numerical score). This score can be seen as a subjective expression of the quality of a wine (Oczkowski, 2016). Price, on the other hand, may be seen as a more objective reflection of a wine's quality, determined by several factors not influenced by the reviewer, but rather by other quality measures such as the growing season average temperature and rainfall (Oczkowski, 2016). Even though the relationship between price and quality is not always one-to-one (Goldstein et al., 2008), price may nevertheless give an accessible, rough approximation of quality, in addition to the more subjective expert rating. A wine review may reflect these aspects of quality of a wine in a number of ways: more expensive wines may be described using more words on average, for example (Ramirez, 2010).

The studies reviewed above exemplify a rich research tradition using statistical analysis of wine review corpora. However, there are few studies that have applied natural language processing (NLP) techniques to such data. To date, the research that has used NLP or data mining techniques of wine data have focused on machine learning applied to data containing information about the chemical components of wine (for example (Urtubia et al., 2007; Cortez et al., 2009)).

A previous study of ours did use NLP methods (Hendrickx et al., 2016), and asked whether wine experts use consistent terminology to describe wine, and if by consequence expert reviews contain enough information to automatically predict wine color, grape type, and country of origin by means of supervised classification techniques. In this paper, we take this work forward and examine *how* experts use consistent terminology to describe wine, and investigate the contribution of different types of information, viz. lexical, semantic and terminological feature groups. Our research differs from previous studies (such as (Brochet and Dubourdieu, 2001)) where corpus-based statistical analyses were used to examine wine language; whereas we train machine learning algorithms on a large corpus of expert wine reviews to automatically predict various characteristics. In addition, we investigate the relationship between subjective and objective wine quality indicators and basic review properties, i.e., the length of the review and average length of the words used in the review.

3. Corpus Description

We collected a corpus of wine reviews from `http://www.winemag.com/`, containing in total 76,410 unique reviews from 33 experts. These wine reviews are combined with additional information about the wines such as the name of the producer, production year, alcohol percentage, color, grape type(s), origin, and rating by the expert. These ratings vary between 80 and 100. The reviews are rather short, on average 39 (untokenized) words per review, and often these reviews combine a sensory description with some additional information, such as the producer or region. The following is an example of a review about an red wine produced in Italy in 2009 with a price of 45$ and which was rated 91 out of 100:

> Cantina del Pino makes some of the finest Barbaresco available today. This shows a succulent quality, with aromas of smoked bacon, wild berries and forest underbrush. Savory and sophisticated, this has loads of personality.

Take, for example, another review of a Spanish red wine with a low rating of 80 and a price of 17$:

> Best on the nose, but sharp and narrow as can be on the palate. Cranberry and sour cherry flavors dominate, while the finish is astringent. No amount of swirling and saving is going help it much.

Note that not all reviews had all metadata fields filled. In our experiments we only use those reviews for which we had non-zero values for the class to be predicted. So, for example the classification experiments excluded the 5,308 reviews of wines where the color was unknown.

4. Classification

We aim to study the usefulness of domain-specific terminology as feature representation for predicting wine properties on the sole basis of the wine review text. We experiment with different feature representations and we compare the terminology features against lexical bag-of-words features, and semantic word embedding features. To operationalize the task of automatically detecting objective wine characteristics, we build supervised machine learning systems and aim at predicting three wine characteristics: color, grape variety, and country of origin.

4.1. Experimental setup

The corpus was randomly split into a training (80%) and test (20%) partition. As evaluation measures, we report averaged micro F-scores on the held-out test set.

To predict the color of a wine, we limit ourselves to three categories: *white*, *red* and *rosé*. In order to have a better

understanding of the predictive power of this classifier, we removed color adjectives referring to the three color classes from the review text. To train a system predicting grape variety, we selected wines produced from a single grape with at least 200 reviews in the training set and removed all grape blends. We also merged grape names referring to the same grape (e.g., *Pinot Gris* and *Pinot Grigio*), which resulted in a total of 28 grape varieties to be predicted. Much variation can be seen in the number of training instances per grape, ranging from 5,706 reviews for *chardonnay* to 222 reviews for *carmenère*. The third classifier aims at predicting among 47 different countries of origin. Again, the class distribution is unbalanced, with some countries represented very well (e.g., US: 25,104 reviews, Italy: 9,912 reviews, France: 8,568 reviews) to countries only occurring once (Tunisia, South Korea, Montenegro, India) in the training set.

All wine reviews were linguistically preprocessed by means of the Stanford toolkit (Manning et al., 2014) involving tokenization, lemmatization and Part-of-Speech tagging. From the preprocessed review text, three different feature types were extracted to model the three classification tasks: lexical, semantic and terminology features.

4.1.1. Lexical features

We extracted a list of bag-of-words (BoW) unigram features from the review text containing lowercased lemmas. These BoW features were filtered on Part-of-Speech category to filter out function words and only keep content words (nouns, adjectives, verbs, and adverbs). The BoW features were incorporated as binary features, meaning that each BoW is a separate feature, which is assigned the value "1" if it occurs in the respective wine review, and "0" otherwise.

4.1.2. Semantic features

In order to reduce data sparsity, we also created word embeddings from the training reviews by means of Word2Vec (Mikolov et al., 2013). After traing, for any input word Word2Vec is able to produce a word vector containing distributional information, i.e., information about the surrounding lexical contexts in which the word occurs.

cluster size	Accuracy
100 clusters	92.398%
200 clusters	94.674%
300 clusters	**95.843**%
500 clusters	95.018%
1000 clusters	95.391%
2000 clusters	95.177%

Table 1: Cross-validation accuracy for a varying number of cluster sizes.

Word2Vec was run with standard settings, that is the BoW model with a context size of 8, and a word vector dimensionality of 200 features. To group word vectors for words that share common contexts in the wine reviews, and thus are located in close proximity in the vector space, we clustered the obtained word vectors using a K-means clustering algorithm. We then encoded the resulting clusters as binary features that were activated if the review text contained a word occurring in the respective clusters. To decide on the desired number of output clusters, we performed 10-fold cross-validation experiments on the training data with a varying number of cluster features (100, 200, 300, 500, 1000 and 2000 clusters). Table 1 shows the accuracy for the cross-validation experiments with varying cluster sizes. A manual inspection of the resulting clusters revealed that the clusters indeed contain semantically related terms. This is illustrated by cluster 82, which contains many terms referring to floral and other related aromas:

> abundant, acacia, aromatic, bee's, clover, dandelion, delicate, enticing, floral, flower, foremost, fragrant, freesia, fresh-cut, freshly, fuzz, garden, jasmine, light-weight, lilac, musk, oils, peony, petroleum, pretty, roses, rosewater, subtle, talcum, wax, wisp, wispy

4.1.3. Terminology features

As a third feature group, we extracted domain-specific terms from the wine review corpus. Terms are linguistically motivated units that refer to concepts within a given domain. The wine-specific terms were extracted by means of TExSIS (Macken et al., 2013), a hybrid terminology extraction tool. In a first step, linguistic preprocessing (Van de Kauter et al., 2013) is run on the wine corpus to perform tokenization, lemmatization, Part-of-Speech tagging, chunking, and named entity recognition. Subsequently, TExSIS makes use of this linguistic information to generate syntactically valid candidate terms. In a final step, statistical filters such as Termhood and C-value are applied to generate the list of single (e.g. *flavor, cherry, ripe, spice, finish*) and multi-word (e.g. *cherry fruit, berry flavors, firm tannins, black currant*) terms.

The underlying idea of the *termhood* filtering is that domain-specific terms (e.g. *tannin, nose*) have much higher relative frequencies in the domain-specific wine corpus than in a standard corpus of English, the Web 1T 5-gram v1 corpus[1] in this case. An inspection of the top-10 terms extracted by TExSIS with the highest termhood scores (i.e. *flavors, tannins, aromas, wine, acidity, fruit, palate, finish, off-dry, cherry*) reveals that these terms indeed belong to specialized wine vocabulary. Although some of these terms also occur in common language, they are much more frequent in the wine corpus than in a general background corpus.

The second statistical filter, the *C-value* filter (Frantzi and Ananiadou, 1999), checks the degree of cohesiveness inside multi-word terms. The C-value metric aims at handling the extraction of nested terms by examining the frequencies of a term used as part of a longer term. Examples of TExSIS multi-word terms with high C-value scores are: *black cherry, fruit flavor, Cabernet Sauvignon, pinot noir, tropical fruit, crisp acidity, dark chocolate* and *smoky oak*.

The resulting terms were again incorporated as binary features in the feature vector. For each review, we extracted

[1] https://catalog.ldc.upenn.edu/ldc2006t13

the list of unigrams, bigrams, trigrams and fourgrams, and set the binary features to '1' if the respective term occurs in one of the n-gram lists. In order to have a comparable amount of bag-of-words and terminological features, we included the same number of features for both feature groups. For the terminological features, we therefore considered the 15,357 most domain-specific terms, sorted by termhood.

4.1.4. Machine learning algorithm

As a classification algorithm, we used Support Vector Machines as implemented in the LIBSVM toolkit (Chang and Lin, 2011). We ran LIBSVM in three different settings: (1) RBF Kernel with standard settings, (2) RBF Kernel with optimized settings (by means of a Grid search, performed on a sample (5,000 randomized instances) of the training data for each classification task, and (3) Linear kernel with standard settings.

When building the feature vectors, we removed reviews where the respective category label was missing, resulting in a training and test set of varying size per classification task. Table 2 lists the number of instances per training and test for each task, as well as the number of categories to be predicted, which gives a good indication of the classification difficulty.

classification task	training	test	categories
colour	56,893	14,209	3
grape type	39,900	9,976	28
country	61,128	15,282	47

Table 2: Size of the training and test sets and number of categories to be predicted for each classification task.

4.2. Classification results and discussion

Table 3 lists the results for the LIBSVM Linear Kernel and optimized RBF Kernel for a varying feature vector. We show experimental results for the different feature groups in isolation and for a combination of all three information sources.

The results of the different feature representations in isolation show that for the grape variety and country classification experiments, the domain-specific terminology features selected by TExSIS outperform the BoW and Word2Vec features. For color however, we see that a simple BoW gives the best performance. In addition, it appears that the choice of SVM kernel does not have a huge effect, as each column shows similar tendencies. When we combine the three feature representations, we again see that this is beneficial for the performance on the grape variety and country classification experiments, but not for the color experiments where BoW features perform best.

We also investigated which domain-specific terms were most informative for the *country* and *variety* classification tasks. To this end, we calculated Information Gain (IG) weighting, which measures for an individual feature how much information it contributes to predicting the correct class label. This was done by computing a probability-weighted average of the informativeness of the different values of the feature, with the IG implementation provided by Timbl (Daelemans et al., 2009). We contrasted an analysis of the best performing terminology features with the most informative lexical descriptors from the bag-of-words feature set.

Setup	RBF opt	Lin Kernel
	Color	
BoW	96.75%	96.59%
Word2Vec	96.31%	96.18%
TExSIS	93.49%	91.66%
All features	96.09%	95.29%
	Grape Variety	
BoW	42.10%	48.28%
Word2Vec	57.39%	56.46%
TExSIS	72.53%	72.77%
All features	76.16%	76.61%
	Country	
BoW	66.50%	60.32%
Word2Vec	69.17%	68.27%
TExSIS	78.67%	79.06%
All features	82.27%	82.84%

Table 3: Averaged micro F-scores per category for the optimised RBF (*RBF opt*) and Linear (*Lin Kernel*) kernels.

Table 4 lists the top-25 most informative terminology (TExSIS) and bag-of-words (BoW) features for both classification tasks.

Country TExSIS	Variety TExSIS	Country BoW	Variety BoW
fruit	aromas	aroma	aroma
it	acidity	palate	tannin
aromas	cherry	acidity	acidity
finish	tannins	cherry	dry
palate	palate	tannin	palate
acidity	ripe	note	note
tannins	dry	ripe	ripe
has	has	spice	spice
cherry	dried-herb	drink	drink
ripe	oak	dry	oak
drink	drink	nose	sweet
notes	notes	black	not
spice	rich	berry	rich
nose	black	not	black
black	spice	sweet	blackberry
fresh	soft	plum	show
sweet	very	fresh	berry
rich	nose	show	now
berry	blackberry	rich	good
dry	shows	red	soft
red	fresh	oak	very
plum	berry	soft	nose
soft	red	good	year
oak	crisp	now	fresh
very	vanilla	blackberry	red

Table 4: Top-25 most informative terminology and bag-of-words features sorted by descending Information Gain scores for the *country* and *grape variety* tasks.

A first observation is that the Information Gain analysis re-

veals a big overlap between the most informative descriptors (1) for the two classification tasks, i.e. *country* and *variety* (overlap of 20 TExSIS and 23 BoW features) and (2) for the two different features groups, i.e., terminology and bag-of-words features (overlap of 20 for *country* and 22 for *variety*). Second, the most informative descriptors do not contain explicit mentions of the class labels (i.e., grape types, countries), but rather contains odour (e.g., *aroma*), flavour (e.g., *sweet, tannin*) and sight (e.g., *red*) descriptors. In addition, they contain source-based (e.g., *plum, oak*), as well as evaluative terms (e.g., *good*), verbs related to the wine tasting process (e.g., *finish, drink*), and more general vocabulary (e.g., *show*).

5. Correlation between price, score and text characteristics

Another aspect we were interested in was how strong the relationship was between basic level properties of the review (i.e., the length of the review in number of characters and the average length of words used), and the objective price of the wine and subjective rating given by the reviewer. To estimate this relationship, two regression models were run on the data: one predicting the price from word length and review length, and one predicting review score from word length and review length.

Average word length per review was established by dividing the amount of characters by the number of words. Reviews with an unknown score or price (standardized to dollars per 0.75L) were removed from the data set, resulting in 67,006 remaining reviews. Analyses were performed using R (R Core Team, 2013). The means and standard deviations of the different review features (price, score, average word length, and review length) can be found in Table 5.

feature	mean	SD
price	33.3	47.9
rating	87.6	3.2
review length	235.7	71.2
word length	6.0	0.4

Table 5: Means and standard deviations for the review features price in dollars, review score in points, and review length and average word length measured in characters.

	Estimate β	SE	t-value	p
Intercept	-12.98	2.74	-4.7	$< .001$
Av word length	2.61	0.45	5.8	$< .001$
Review length	.13	.003	50.9	$< .001$

Table 6: Predicting price from average word length and review length.

The regression model for review price and rating for all 67,006 reviews which had non-zero values are shown in Table 6 and Table 7 respectively. Review length and average word length were significantly related to the price of the wine, in such a way that for every additional character in a review, the wine was on average 13 cents more expensive, and the use of longer words by one character predicted an extra 2.6 dollars on top of the wine price, on average. Similarly, the review score was significantly correlated with review length and word length. For each additional character in the review, the score of the wine increased by .79 points. Average word length in a review had less influence, as the use of one-character longer words increased the score by only .025 points.

	Estimate β	SE	t-value	p
Intercept	7.76	.164	479.3	$< .001$
Av word length	.791	.027	29.5	$< .001$
Review length	.025	.002	149.6	$< .001$

Table 7: Predicting review rating score from average word length and review length.

6. Conclusion

This paper describes a set of classification and regression experiments aimed at predicting wine characteristics based on the review text. The results show that (1) wine experts indeed share a common vocabulary, making it possible to predict the color, grape variety and country of origin of the wine to a reasonable extent, and that (2) terminological features outperform bag of word features and semantic features when used in isolation. In addition, review length and average word length were shown to be significantly related to review price and rating.

In sum, this study shows that the language of wine reviews is richly informative (contra previous claims), and demonstrates the important role of NLP methods to address core questions about the limits and possibilities of language more generally.

7. Bibliographical References

Brochet, F. and Dubourdieu, D. (2001). Wine descriptive language supports cognitive specificity of chemical senses. *Brain and Language*, 77:187–196.

Caballero, R. (2007). Manner-of-motion verbs in wine description. *Journal of Pragmatics*, 39(12):2095–2114.

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27:1–27:27. ISSN: 2157-6904.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553.

Croijmans, I. and Majid, A. (2016). Not all flavor expertise is equal: The language of wine and coffee experts. *PLoS ONE*, 11(6):e0155845.

Daelemans, W., Zavrel, J., Sloot, K., and Van Den Bosch, A. (2009). TiMBL: Tilburg Memory-Based Learner, version 6.2, Reference Guide. ILK Technical Report 09-01.

Frantzi, K. and Ananiadou, S. (1999). The C-value / NC-value domain independent method for multi-word term extraction. *Journal of Natural Language Processing*, 6:145–179.

Gawel, R. (1997). The use of language by trained and untrained experienced wine tasters. *Journal of Sensory Studies*, 12(4):267–284.

Goldstein, R., Almenberg, J., Dreber, A., Emerson, J. W., Herschkowitsch, A., and Katz, J. (2008). Do more expensive wines taste better? evidence from a large sample of blind tastings. *Journal of Wine Economics*, 3(1):1–9.

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*, pages 306–312, Berlin, Germany, August. Association for Computational Linguistics.

Levinson, S. C. and Majid, A. (2014). Differential ineffability and the senses. *Mind & Language*, 29(4):407–427.

Macken, L., Lefever, E., and Hoste, V. (2013). TExSIS: Bilingual Terminology Extraction from Parallel Corpora Using Chunk-based Alignment. *Terminology*, 19(1):1–30.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and Mc-Closky, D. (2014). The stanford corenlp natural lan- guage processing toolkit. In *Proceedings of the Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60. Association for Computational Linguistics.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 26:3111–3119.

Oczkowski, E. (2016). Identifying the effects of objective and subjective quality on wine prices. *Journal of Wine Economics*, 11(2):249–260.

Quandt, R. E. (2007). On wine bullshit: some new software? *Journal of Wine Economics*, 2(02):129–135.

R Core Team, (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramirez, C. (2010). Do Tasting Notes Add Value? Evidence from Napa Wines. *Journal of Wine Economics*, 5(1):143–163.

Rouby, C., Schaal, B., Dubois, D., Gervais, R., and Holley, A. (2002). *Olfaction, taste, and cognition*. Cambridge University Press.

Solomon, G. (1990). Psychology of Novice and Expert Wine Talk. *The American Journal of Psychology*, 103(4):495–517.

Solomon, G. (1997). Conceptual Change and Wine Expertise. *The Journal of the Learning Sciences*, 6(1):41–60.

Sperber, D. (1975). *Rethinking symbolism*. Number 11. CUP Archive.

Suárez Toste, E. (2007). Metaphor inside the wine cellar: On the ubiquity of personification schemas in winespeak. *Metaphorik. de*, 12(1):53–64.

Urtubia, A., Pérez-Correa, J. R., Soto, A., and Pszczolkowski, P. (2007). Using data mining techniques to predict industrial wine problem fermentations. *Food Control*, 18(12):1512–1517.

Van de Kauter, M., Coorman, G., Lefever, E., Desmet, B., Macken, L., and Hoste, V. (2013). LeTs Preprocess: The multilingual LT3 linguistic preprocessing toolkit. *Computational Linguistics in the Netherlands Journal*, 3:103–120.