beverages

MDPI

Article

Using Latent Semantic Analysis to Investigate Wine Sensory Profiles—Application in Swedish Solaris Wines

Gonzalo Garrido-Bañuelos [1,2,*], Mpho Mafata [1,3] and Astrid Buica [1]

1 School for Data Science and Computational Thinking, Stellenbosch University, Stellenbosch 7599, South Africa; mafata@sun.ac.za (M.M.); astrid.buica@gmail.com (A.B.)
2 Innovate Solutions, D18 X3X7 Dublin, Ireland
3 Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch 7599, South Africa
* Correspondence: gonzalo@innovatesolutions.ie

Abstract: Online text is a source of data in many fields, but it is yet to be explored by sensory scientists. The present work aimed to explore the suitability of using a bibliometric methodology such as Latent Semantic Analysis (LSA) to understand and define wine sensory spaces. Data were also explored by the more conventional Multiple Correspondence Analysis (MCA). The present work shows the potential use of LSA in sensory science; the first part of the study investigates the sensory profile of Swedish Solaris wines, while the second part focuses on understanding their fit with two international monovarietal white wines (Albariño and Chenin Blanc). The results show that the majority of Swedish Solaris wines could be associated with two different styles (LSA topics). However, there is no evidence of a cultivar typicality, as when comparing the Solaris wines with Albariño and Chenin Blanc, they shared features with both cultivars. Chenin Blanc was also found to be associated with different styles. In contrast, Albariño wines showed to have more unique features as the majority were associated with a single LSA topic.

Keywords: Solaris wine; sensory space; text data; LSA; MCA; data science

check for updates

Citation: Garrido-Bañuelos, G.; Mafata, M.; Buica, A. Using Latent Semantic Analysis to Investigate Wine Sensory Profiles—Application in Swedish Solaris Wines. *Beverages* **2024**, *10*, 120. https://doi.org/10.3390/beverages10040120

Academic Editor: Enrique Durán-Guerrero

Received: 17 October 2024
Revised: 11 November 2024
Accepted: 25 November 2024
Published: 3 December 2024

1. Introduction

The international wine panorama has evolved in recent decades. This evolution was partly due to emerging wine-producing countries in regions with more extreme climatic conditions, such as Brazil and Thailand [1–3], which have tropical climates, and the Nordic regions, such as Denmark and Sweden [4–6]. Nordic countries have been favored by the continuous rise in global temperatures, leading to climatic conditions more suitable for vine growing. Even so, most of the wines produced are still made from frost-resistant cultivars, such as the *Vitis vinifera* hybrids Solaris (white cultivar) and Rondo (red cultivar) [7,8].

Originating in Germany in 1975, the Solaris grape is the interspecific hybrid result of the cross between Merzling (Seyve-Villard × Riesling × Pinot 56 Gris) with another crossing Gm6493 (Zarya Seva × Muscat Ottonel) [8]. Despite not being as well known internationally as other white varieties, Solaris wines are already being produced in some emerging wine-producing countries, but also in more traditional wine regions such as Italy and Germany [9]. The relevance of Solaris wines and their representation on the international wine panorama can be shown based on the number of scanned Solaris wines on Vivino, a consumer-based wine platform [10]. By September 2021, a total of 241 monovarietal Solaris wines from fourteen different countries have been scanned by consumers. Denmark is the country with the largest number of Solaris wines (39), followed by Germany (31) and Switzerland (30), with Sweden having the fifth largest number (28). Solaris wines from countries such as Italy (12), France (1), and the United States (1) were also found on Vivino. This information is, however, based on consumers' scans, which might be subjected to many different influences (i.e., wine availability and cultural habits) and may not represent complete data on global Solaris wine production. Nonetheless, it shows that its production and consumer interest are not limited to Nordic regions.

Even though it is possible to say that several Solaris "styles" can be made, published information about the Solaris sensory space and potential styles is very limited. In 2014, Swedish Solaris wines were characterized by "citrus" and "floral" aromas [1–3]. A more recent publication proposed that Swedish Solaris wines could fall into three aroma categories: green, fruity, and oaky. The results obtained by Garrido-Bañuelos et al. (2020) [11] were similar to those in a previous study on Danish Solaris wines conducted by Liu et al. (2015) [5], who found that Solaris wines could be perceived as either floral or fruity or described with attributes such as "wood" or "rooibos/smoke". The study by Garrido-Bañuelos et al. (2020) [11] included a sensory evaluation of typicality as perceived by Swedish winemakers, and the results show a clear lack of agreement. These results indicate that the Solaris sensory space still needs to be further explored and defined, and a stronger concept related to the perception of Solaris wines' typicality and styles still needs to be developed.

The sensory space of a product is a complex concept not only defined by intrinsic properties (i.e., color, texture, aroma, and taste) but also by extrinsic factors (e.g., the levels of acceptance and preferences of the tasters) [12]. Investigating the sensory space of a product depends on whether or not it has been previously defined and/or tested. If it has been properly defined, the conceptual (ideas) and perceptual (tasting) sensory aspects can be evaluated. In wine, the level of experience of an assessor may have a bigger influence on conceptual perception than perceptual perception [13].

From an experimental perspective, the process of characterizing a wine's sensory space can be divided into three phases, as stated by Barbe et al. (2021) [14]: the identification of a sensory concept, the perceptual evaluation of the sensory concept, and sensory space description. Mafata et al. (2020) [15] investigated the typicality (cognitive) and the sensory space (perceptual) of South African old vine Chenin Blanc wines; however, the existence of a unique sensory space could not be demonstrated mainly due to a lack of perceptual consensus among wine industry professionals [15].

Text data, in combination with different data strategies, have become a potential tool to unravel inherent sensory features and have been utilized to perform a diversity of sensory studies, which vary from understanding the cross-cultural impact on the term "floral" [16] to creating and defining the lexicon of existing products such as rum [17], whisky [18], and monovarietal apple juices [19]. Da Silva et al. (2019) [19] used text and data mining strategies to generate an aroma wheel based on different peer review publications; this aroma wheel was then used to carry out the sensory evaluation of different juice samples. Text data have also been used as a potential source to understand the profile of a specific market (i.e., Swedish beers) or the properties of specific raw material within a product, such as hops [20]. Garrido-Bañuelos et al. (2021) [20] used a dataset of a similar nature to the one used in the current study. The results show how text data can help people understand sensory similarities between different hop varieties, such as Amarillo, Chinook, and Magnum, characterized by the presence of "spicy" and "dried fruit" descriptors. For wines, Valente et al. [21] used text data extracted from the annually published John Platter Wine Guide to South African Wines ("Platter's Wine Guide", [22]) to model the sensory space of South African Sauvignon Blanc and Chenin Blanc wines; the authors coupled text mining with machine learning algorithms to explore the association between specific sensory descriptors and Chenin Blanc wine styles. Different Natural Language Processing (NLP) techniques have also been used to explore the lexicon extracted from online reviews by wine experts and their similarity to the vocabulary used to train wine experts [23,24]. Both studies based their analyses on wines reviews with scores higher than 80 points (on a scale from 0 to 100), obviating a potential lexicon which could be associated with wines with a generally lower perceived quality. NLP has also been applied to explore the sensory drivers of "outstanding" and "extraordinary" wines and the consistency of major WineSpectator reviewers to describe these wines [25]. Within the sensory space of a product, sensory attributes can be both positive and negative drivers of its perceived quality, as shown for Pinotage and Chenin Blanc wines [26].

From a methodological perspective, if a sensory concept is well defined, a closed-ended sensory experiment can be followed by supervised statistical methods to test the concept. However, if a concept is ill defined or unknown, open-ended exploratory methodologies followed by unsupervised statistical methods should be used [27]. Given the complex definition of sensory space, mixed-methodology approaches can be used where a concept is ill defined, or where a concept is well defined but there are no prototypes, or the borders of typicality between styles are investigated [28]. Hence, tailoring data strategies is essential to obtaining the best outcomes. The nature of the data and a clearly defined research question will help guide the choice of data strategies to be used. In sensory science and consumer science, data strategies can range from classical statistics and multivariate analysis on individual datasets to data fusion on multiblock datasets [29–32]. For example, and within the scope of this work, Correspondence Analysis (CA) and Multiple Correspondence Analysis (MCA) have been frequently used to analyze qualitative data [33–36]. Recent studies are directing their work to the use of machine learning strategies [21,37–39] and more complex mathematical models which can help our understanding of the sensory space complexity [40,41].

Latent Semantic Analysis (LSA) is a methodology commonly used in computer science for language processing and topic analysis, with applications in various fields such as linguistics, bibliometrics, and cognitive science [42]. It has also been used as a tool to investigate how customers' emotions can have a positive or a negative effect on product reviews from Amazon [43]. In short, LSA is a statistical methodology used to compute semantic concepts from latent semantic information contained in large text documents [44]. These concepts are defined by the relationship between a set of documents and terms found in the text. LSA operates on sparse data, similarly to MCA; however, the purpose of LSA is to generate a single decomposition clustering around specific semantic spaces called topics [42,45]. LSA results provide the number of terms and number of documents associated with each topic. The outcome also provides information about term-to-term, document-to-document, and term-to-document relationships associated with the topics. Topics are formed depending on how often the terms are associated with one another. Nonetheless, LSA does not provide information about the relationships between the proposed topics. The recent integration of LSA algorithms into the XLSTAT software starting with version 2018.1 also facilitated its inclusion in the present study.

Despite LSA's multiple applications in text and data mining, its application in sensory science has not been explored by using attributes (i.e., terms) and products (i.e., documents) to explore sensory space and potential styles/categories (i.e., "topics"). The source of our dataset is the Systembolaget website. Systembolaget is a governmental alcohol retail monopoly established in Sweden during the mid-1800s with the aim to reduce alcohol-related problems. Currently, Systembolaget provides information about each individual product available in the Swedish market, which includes a short sensory description for each of them. The present work aimed to use data science to bring new insights, which can contribute to the definition of Swedish Solaris wines' sensory space, and to understand this cultivar's fit on the international wine panorama by comparing Solaris with two international grape cultivars (Chenin Blanc and Albariño). The selection of these two international wine cultivars was based on findings from Garrido-Bañuelos et al. (2020) [11] that showed how during a blind evaluation of different international white wines, Albariño and Chenin Blanc wines were considered to be typical Swedish Solaris by Swedish winemakers. Moreover, this work explored the suitability of LSA as a potential tool to understand sensory insights from text data in comparison to other, more conventional data analysis tools, such as MCA.

2. Materials and Methods

2.1. Data Collection

All data for the current study were extracted between August and September 2021 from the website of the Swedish National alcohol retail monopoly, Systembolaget [46]. Data were captured from 36 Swedish Solaris monovarietal wines (SOL), 16 Chenin Blanc (CB), and 19 Albariño (Alb) wines. Wines were labeled in short-hand according to cultivar and country of origin (i.e., Sweden—SW; Spain—SP; Portugal—PT; France—FR; South Africa—SA; Australia—AUS; and Uruguay—U). The data capturing, cleaning, and standardization procedures were as described by Garrido-Bañuelos et al. [20]. In short, the list of sensory attributes provided for each wine was captured in Swedish and then translated to English with the aid of Google translate; the text was tabulated, and words with hedonic and intensity connotations were excluded. A nominal table was then built based on the presence/absence of all sensory attributes. A similar approach was used by Mafata et al. [47]. Due to a small sample size, several sensory attributes were only present in a low number of wines. Therefore, attribute consolidation was not performed.

2.2. Data Analysis

A schematic workflow summarizing the data strategies followed for this work can be found in Figure 1. LSA and MCA were performed using the same raw data, consisting of a nominal table of sensory descriptors (categories of presence—1; absence—0) and samples (wines). The LSA was carried out using hard clustering. The optimal number of topics was chosen based on guided inspection and a cumulative percentage variance (%EV) of approximately 70, which has been shown to allow for optimal results in exploratory analyses [48,49]. A proximity matrix using the Euclidean distance, based on co-occurrence, was generated from the presence of the different sensory descriptors for each LSA topic [44]. Note that this is because LSA is based on term associations; hence, the document (wine) associations were not used to generate the proximity matrix. Multidimensional Scaling (MDS) was performed on the proximity matrix to visualize the relationships between topics. MCA was performed based on adjusted inertia [20]. Agglomerative Hierarchical Clustering (AHC) was carried out following Ward's method for both MCA and MDS dimensions. AHC was performed on the first eight dimensions (cumulative variation $\geq$ 70%) of the MCA scores [49] and the first two dimensions of the MDS (approximately 0.20 Kruskal's index). All data analyses were performed with XLSTAT 2021 (Addinsoft (2021), New York, NY, USA).

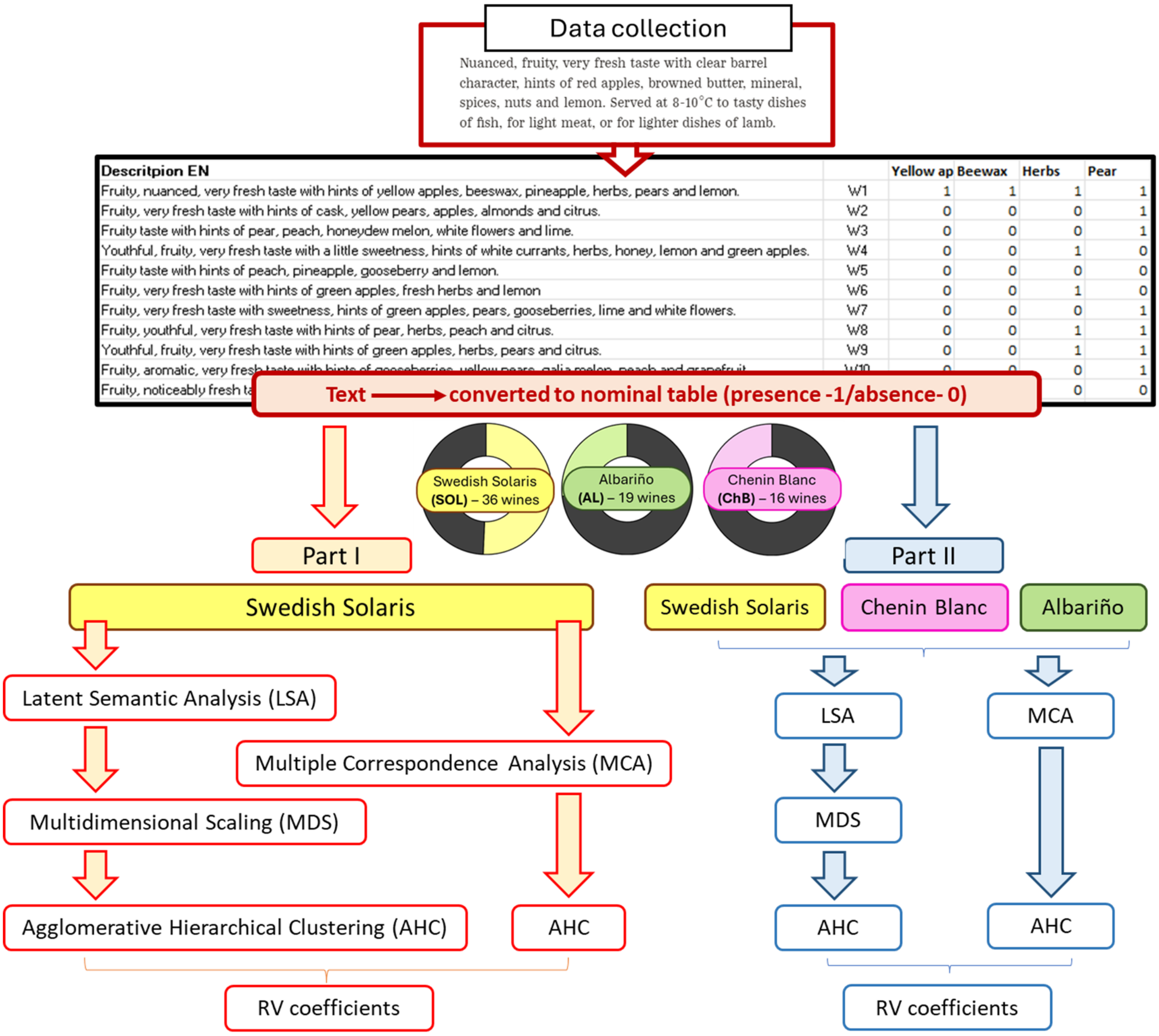

Figure 1. Schematic workflow of data strategy used to understand Swedish Solaris sensory space and its comparison with two international cultivars, Chenin Blanc and Albariño.

3. Results and Discussion

3.1. Sensory Space of Commercial Swedish Solaris Wines

A total of 41 different aroma descriptors were used to describe the aroma profiles of the 36 Swedish Solaris wines commercially available at the date of the study. The term "pear" was the most frequent (17 citations; 41% wines). At the other end of the scale, the following ten attributes (≈25% of descriptors used) were only cited once: "beeswax", "bergamot", "blackcurrant leaves", "cardamom", "grapefruit", "hazelnut", "orange flower", "orange peel", "watermelon", and "yellow kiwi". Due to the low number of wines and the relatively large number of attributes with a low citation frequency, attribute consolidation was not taken into consideration.

3.1.1. Exploring Potential Wine Styles of Swedish Solaris Wines Using Latent Semantic Analysis

LSA was used to find associations between wines and topics, which were represented by semantic spaces (i.e., potential wine styles). These topics, and their associations with sensory attributes, were explored following a hard clustering approach. As an initial exploratory step, the hard clustering process was initially set over ten topics (as a default setting). The explained variance was 80% for the ten topics, with four topics being associated with a single sample and one topic not being associated with any samples (an unrepresented topic) (Table S1). In order to overcome the sparsity (where no samples are associated with the attributes in any given topic) and underrepresented topics, the selection of the optimal number of topics was based on the %EV. Topic 1 had the largest weight with 22.15%, followed by Topic 2 with 13.67%. From Topic 3 onwards, each topic weight represents less than 10%, and from topic 7 onwards, each topic represents less than 5% of the data (Tables 1 and S1). Therefore, seven topics were chosen based on a cumulative inertia of 68.7%.

Table 1. Table showing Swedish Solaris wine topics and associated terms following LSA with hard clustering.

Topics	Number of Documents (“Wines”)	Number of Documents (“Wines”)	Variability (%)	Documents	Terms (“Sensory Attributes”)
Topic 1	16	5	22.2	W1, W2, W3, W4, W7, W8, W9, W10, W12, W13, W14, W22, W23, W28, W29, W36	Pear, Herbs, Green Apple, Citrus, Honey
Topic 2	6	12	13.7	W17, W18, W19, W20, W25, W26	Oak, Apricot, Butter, Yellow Apples, Nut, Red Apple, Pineapple, Rosehip, Marzipan, Hazelnut, Watermelon, Cardamom
Topic 3	5	4	8.2	W6, W15, W33, W29, W21	Lemon, Yellow Pear, Almond, Galia Melon
Topic 4	3	5	7.6	W34, W27, W11	Lime, Gooseberry, White Flower, Vanilla, Blackcurrant Leaf
Topic 5	3	7	6.3	W5, W30, W24	Peach, Mineral, White Currant, Yellow kiwi, Orange Flower, Grapefruit, Beeswax
Topic 6	2	3	5.6	W32, W16	Orange, Apple, Yellow Plum, Orange Peel
Topic 7	1	3	5.1	W3	Elderberry, Spices, Honeydew melon, Bergamot

Figure 2 shows the document-to-document (i.e., wine-to-wine) association as a correlation matrix based on a degree of similarity with values between 0 and 1. Wines are color-coded based on the topic and, within a topic, the degree of similarity is represented by the gradient tonality. The results show that the largest number of wines (16) were associated with Topic 1, described with five sensory attributes: “pear”, “herbs”, “green apple”, “citrus”, and “honey”. Since the terms “pear” and “herbs” were the most frequently used, it stands to reason that they were the most associated with each other, therefore contributing to the high variability of Topic 1 (Table 1). Therefore, the high variability of Topic 1 is based on both the greatest number of wines and the most frequently cited terms.

Similarly, Topic 2 was characterized by the second highest number of wines (six wines). Additionally, in this case, Topic 2 had the largest number of terms. These included woody and spicy terms such as “oak”, “butter”, “nut”, “hazelnut”, and “cardamom”, as well as fruity aroma terms such as “apricot”, “red apples”, and “pineapple” (Table 1). Hierarchically, Topic 3 had the third largest number of wines, leading to a cumulative amount of 75% of wines (27 out of 36) in the first three topics. Topic 4 to Topic 6 had similar numbers of wines and terms with seemingly non-distinct characteristics.

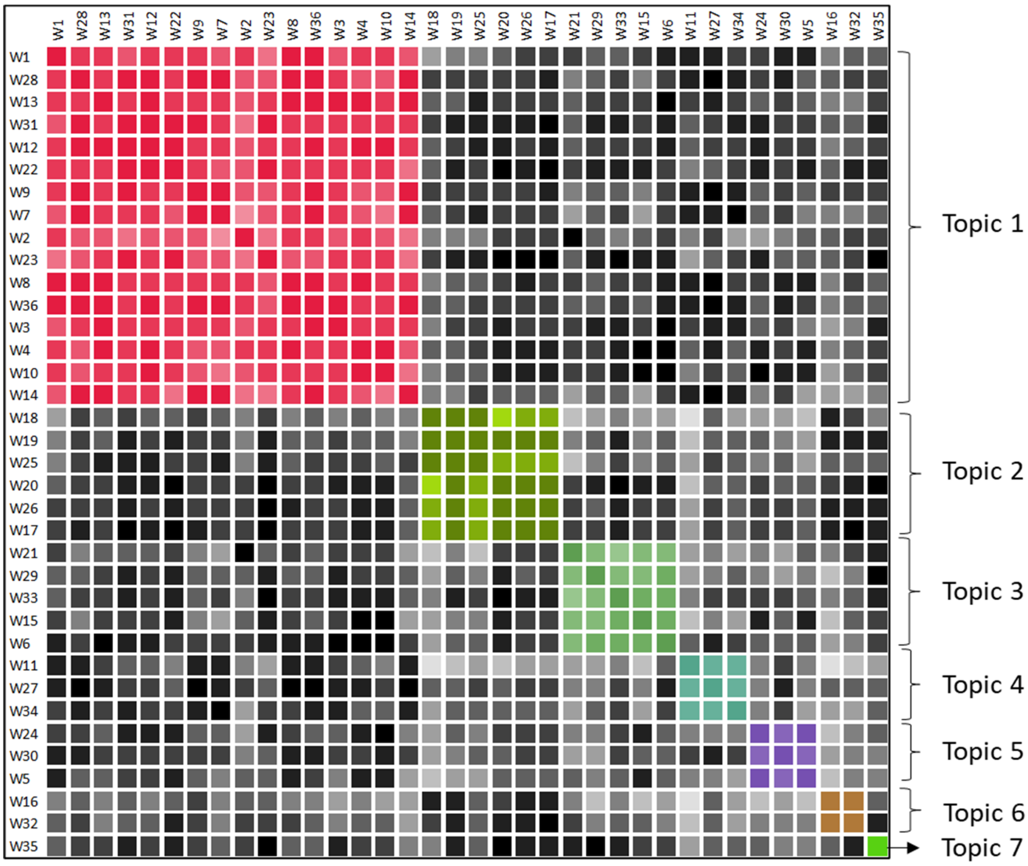

Figure 2. Wine-to-wine correlation matrix built from Latent Semantic Analysis and hard clustering. Wines are associated and color-coded based on specific topics.

The last topic, Topic 7, was characterized by aromas such as "elderberry", "spices", "honeydew melon", and "bergamot", all associated with a single wine (W35). The exceptional element in this wine was that it was the only wine described using the term "bergamot". Although there were ten other terms cited only once, wine W35 was the only one singled out in a topic even though it also had the highly cited term "pear" used in its description. This could be a limitation of using hard clustering and/or LSA on data with a relatively low number of samples. The combination of data sparsity and low density (low number of samples) could lead to misrepresentation of the relationship between samples and between terms.

Therefore, LSA using fuzzy clustering was not considered for this sample set given that at such a low data density (41 attributes and 36 wines), the probabilities of association with every topic can be very similar, making it difficult (fuzzy) to distinguish between topics. Fuzzy approaches are better suited for larger datasets and/or more distinguishable classes [50]. Since the hard clustering approach was taken in this case, some caution must be used when making inferences since the model places a hard border when assigning attributes and documents to a topic. The wines and terms may have very similar associations but can only be assigned to one topic, even if their association to another topic is very similar.

In that regard, the next logical assessment of the LSA is to look at how similar the topics are to one another using the wine and term frequencies from the raw data. A similarity matrix was calculated from the LSA topics (as independent observations) and terms (as dependent variables), with the frequencies based on the number of wines associated with each topic having the given attribute. The similarity matrix was submitted to MDS and AHC to visualize the similarity between the different LSA topics (Figure 3). Figure 3 shows that Topic 1 and Topic 2 are at the farthest branches of the dendrogram and at diagonally opposite quadrants of the MDS plot. This means that these two topics are the most dissimilar. According to AHC, Cluster 1 had Topic 1 and Topic 3 clustered together. This clustering means that the pair has the highest number of samples associated with them and most of the frequently cited terms. Although they are clustered together, their dissimilarity is higher than any two given topics, and given their large Euclidean distance in the MDS, these may be distinct topics. To the far right of the dendrogram, Cluster 2 has Topic 2 and Topic 6 clustered together, and they are more similar than the Cluster 1 pair. Given the high similarity, Topic 6 may be a subset of Topic 2. The minor topics, Topics 3, 4, and 7, with a cumulative variation of 20%, form the most similar cluster (Table S1). A high cophenetic correlation coefficient (0.709) confirmed the reliability and accuracy of the clusters in the dendrogram (Figure 3).

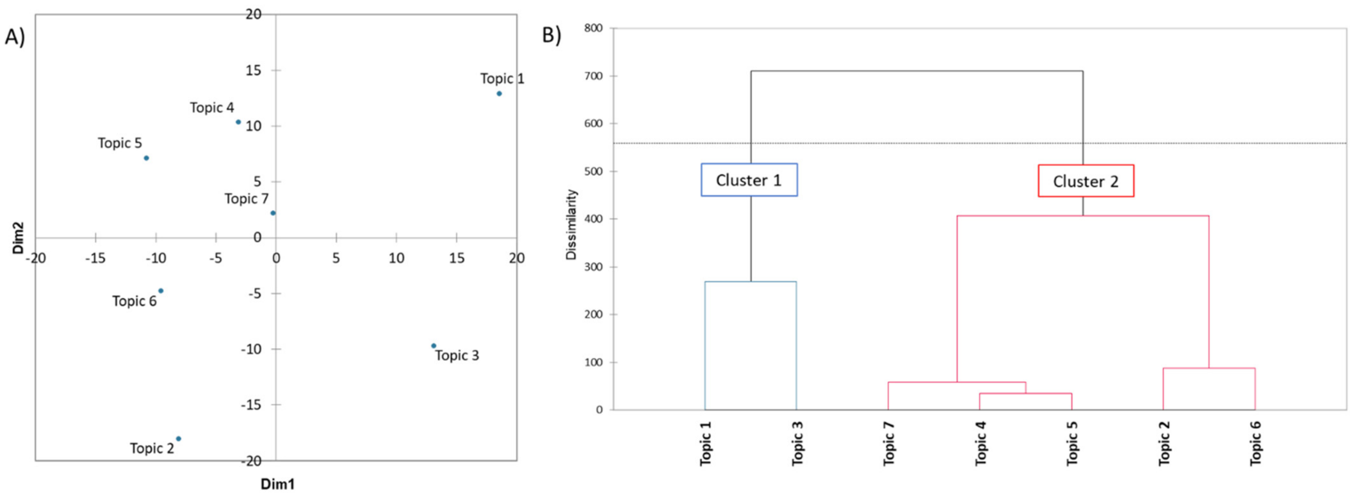

Figure 3. (**A**) Multidimensional Scaling (MDS) and (**B**) Agglomerative Hierarchical Clustering (AHC) of LSA topics of Swedish Solaris wines.

3.1.2. Exploring the Sensory Profiles of Commercial Swedish Solaris Wines Using MCA

The MCA biplot in Figure 4A illustrates the distribution of Swedish Solaris wines and how they corresponded with the sensory descriptors. The MCA showed that the category-0 variables were clustered around the origin, meaning that they had very little contribution to variation and likewise to the distribution of samples [51]. To improve the interpretation and obtain a clearer visualization of the biplot, labels from the category (−0) variables were removed, but the points were still projected. The first two dimensions of the MCA explained about 40% of the variation and showed two patterns, namely, compact clustering on the negative F1 axis and a broad distribution of samples along the positive F1 axis. This clustering pattern was confirmed by the corresponding AHC; the samples grouped into two clusters (Figure 4B), Cluster 1 (compact clustering) and Cluster 2 (broad clustering), and the cophenetic correlation coefficient was 0.645.

Looking at the variable contributions to the first two dimensions, descriptors such as “oak-1”, “nut-1”, “apricot-1”, and “red apples-1” were the major contributors along F1. These sensory attributes described wines from Cluster 2 (Figure 4B), and they were the same terms which fell under LSA Topic 2 (Table 1). Similarly, Cluster 2 (Figure 4B) contains the same wines associated with LSA Topic 2 (Table 1), and the same terms used for the LSA (i.e., “pear-1” and “herbs-1”) correspond with this cluster (Figure 4A). When looking further at the distribution along F2, attributes such as “vanilla-1” and “butter-1” had a large contribution to the negative axis, whereas “rosehips-1” and “orange flower” contributed to the positive axis. These attributes mainly distinguish samples in Cluster 2.

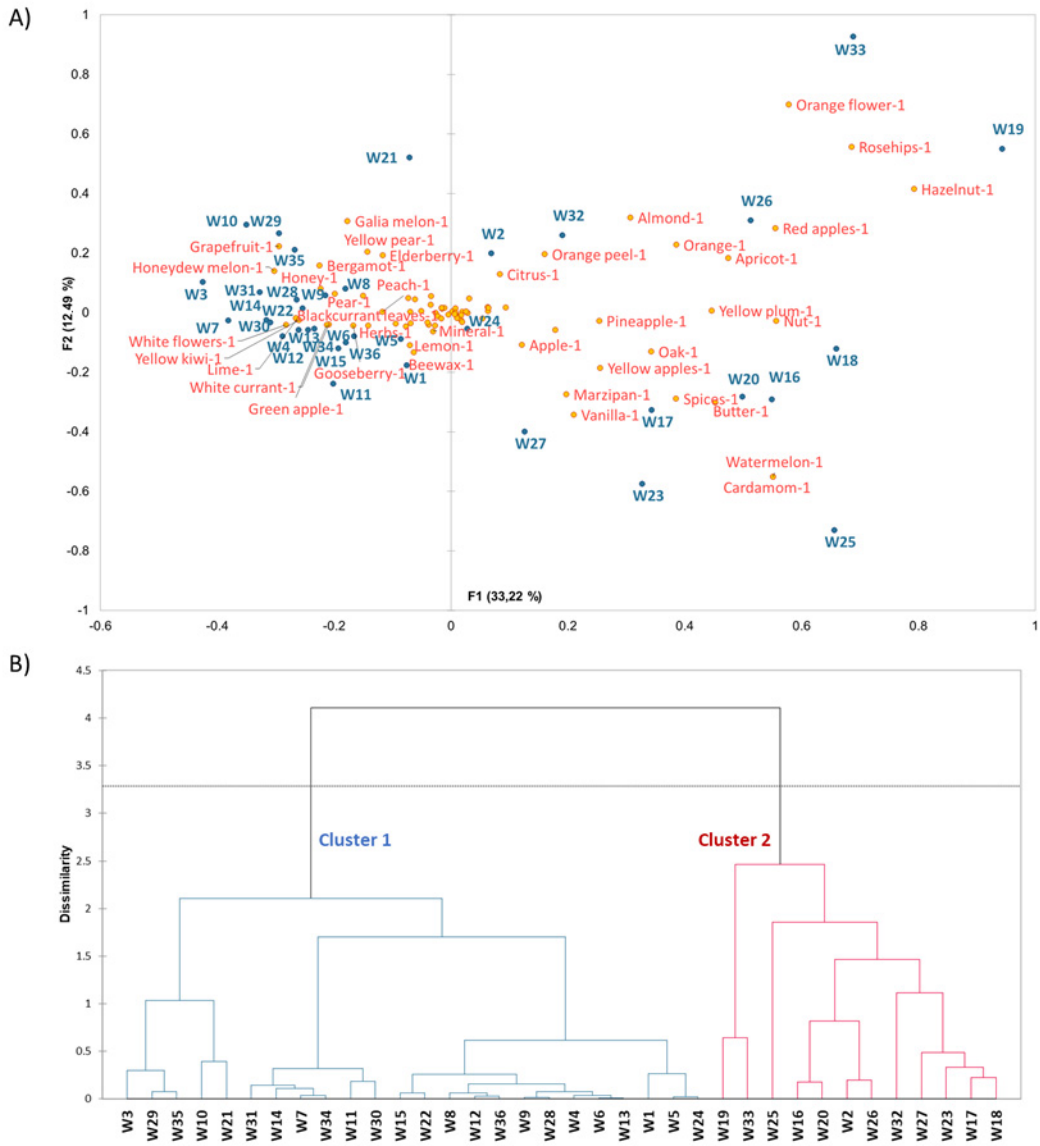

Figure 4. (**A**) Multiple Correspondence Analysis (MCA) biplot projecting Swedish Solaris wines (observations) based on their sensory descriptions and absence/presence (0/1) categories. Only sensory attributes used to describe wines (−1) were included. (**B**) Agglomerative Hierarchical Analysis (AHC) was performed on same data.

3.2. Comparison of Solaris Wines to Monovarietal Chenin Blanc and Albariño Wines

In this section, the Solaris dataset (36 wines) is integrated with monovarietal Chenin Blanc (16 wines) and Albariño (19 wines) wines. These additional cultivars were selected based on the sensory space similarity to Solaris found in a previous study [11]. The addition of these two cultivars resulted in a new dataset consisting of a 71 × 53 (wines × descriptor terms) matrix with twelve new sensory descriptors. Similarly to the Solaris dataset, the terms "pear" and "herbs" were the most cited descriptors. A total of 12 descriptors were mentioned only once, most of which (7) were from the Solaris dataset.

3.2.1. A Latent Semantic Analysis of the Three Cultivars

Using the same strategy used for the Solaris dataset, the optimal number of topics used for LSA was nine based on the approximately cumulative variation of 70% over the nine topics (Table 2). Introducing Albariño and Chenin Blanc impacted the specific sensory attributes associated with each specific topic. The number of terms associated with Topic 1 (12) was larger than that for Swedish Solaris (5) (Table S2). Topic 1 was characterized by the terms “pear” and “herbs”, which were the most frequently cited terms, similar to Solaris Topic 1 (Table 1). Including the two cultivars resulted in the addition of the terms “peach”, “lime”, and “mineral” (Table 2), which were the third, fourth, and fifth most frequently cited terms (Table S2). Topic 2 still had 14 descriptors, the majority of which were associated with Topic 2 for the Solaris dataset, i.e., “oak”, “butter”, “nut”, “apricot”, “red apples”, “cardamom”, “marzipan”, “hazelnut”, and “watermelon”. Although Topic 3 had the third highest variation because it contained some of the top cited terms, it was only associated with one wine. Conversely, Topic 6 had lower variation, although it also contained some top cited terms and was associated with six wines. These are consequences of the LSA prioritizing the associations between terms over documents. Topic five was predominantly associated with terms only cited once, with “citrus” being the only frequently cited term. Similarly, Topics 7 and 8 contained terms cited only once, hence the low variation in these topics.

Table 2. Swedish Solaris, Chenin Blanc, and Albariño wine topics and associated terms.

Topics	Number of Documents (“Wines”)	Number of Terms (“Sensory Attributes”)	Variability (%)	Documents (“Wines”)	Terms (“Sensory Attributes”)
Topic 1	42	12	27.4	Sol-SW1, Sol-SW2, Sol-SW3, Sol-SW7, Sol-SW8, Sol-SW9, Sol-SW10, Sol-SW12, Sol-SW13, Sol-SW14, Sol-SW22, Sol-SW28, Sol-SW31, Sol-SW35, Sol-SW36, ChB-FR1, ChB-SA1, ChB-SA3, ChB-FR3, ChB-SA5, ChB-FR5, ChB-SA6, ChB-FR6, ChB-SA8, ChB-SA10, AL-PT1, AL-SP1, AL-SP2, AL-PT3, AL-U, AL-SP3, AL-PT4, AL-PT5, AL-SP4, AL-PT6, AL-AUS, AL-PT7, AL-PT8, AL-FR, AL-SP6, AL-SP7, AL-SP8	Pear, Herbs, Peach, Mineral, Lime, Yellow Pear, Apple, Beeswax, Gooseberry, White Currant, Smoky, Mineral, Kiwi
Topic 2	12	14	10.4	Sol-SW16, Sol-SW17, Sol-SW18, Sol-SW19, Sol-SW20, Sol-SW23, Sol-SW25, Sol-SW26, ChB-SA2, ChB-FR2, ChB-SA7, AL-SP5	Oak, Nut, Orange, Yellow Plum, Apricot, Butter, Vanilla, Spices, Red Apple, Cardamom, Marzipan, Dried Pineapple, Hazelnut, Watermelon
Topic 3	1	5	6.6	ChB-SA4	Lemon, Pineapple, Grilled Lemon, Macadamia Nut, Nectarine
Topic 4	2	2	5.5	AL-PT2, Sol-SW6	Galia Melon, White Peach
Topic 5	4	5	4.9	Sol-SW27, Sol-SW29, Sol-SW33, ChB-FR4	Citrus, Dried Apricot, Passion Fruit, Apple Blossom, Blackcurrant Leaves
Topic 6	5	3	4.55	Sol-SW4, Sol-SW11, Sol-SW15, Sol-SW30, Sol-SW34,	Green Apple, White Flower, Yellow Kiwi
Topic 7	2	4	4.0	Sol-SW21, Sol-SW24	Honey, Elderberry, Mango, Bergamot
Topic 8	2	3	3.9	Sol-SW5, ChB-SA9	Yellow Apple, Grapefruit, Almond Flower
Topic 9	1	5	3.3	Sol-SW32	Honeydew Melon, Almond, Rosehip, Orange Flower, Orange Peel

Most wines (54 out of 71, representing 76% of the set) were associated with the first two topics (Topic 1—42 wines; Topic 2—12 wines). Topic 1 and Topic 2 represent 27.41% and 10.36% of the variability, respectively. Low variability in sensory data is a known characteristic of these high variability data [28]. Wines from all three cultivars were associated with Topic 1 (Figure 5), predominantly featuring Albariño wines (17; 40%), followed by Solaris (15; 36%), and lastly, Chenin Blanc (10; 24%). Topic 1 included thirteen Solaris wines previously associated with Topic 1 in the Solaris dataset. This topic included the majority of all Albariño wines (17 out of 19) and Chenin Blanc (10 out of 17) wines.

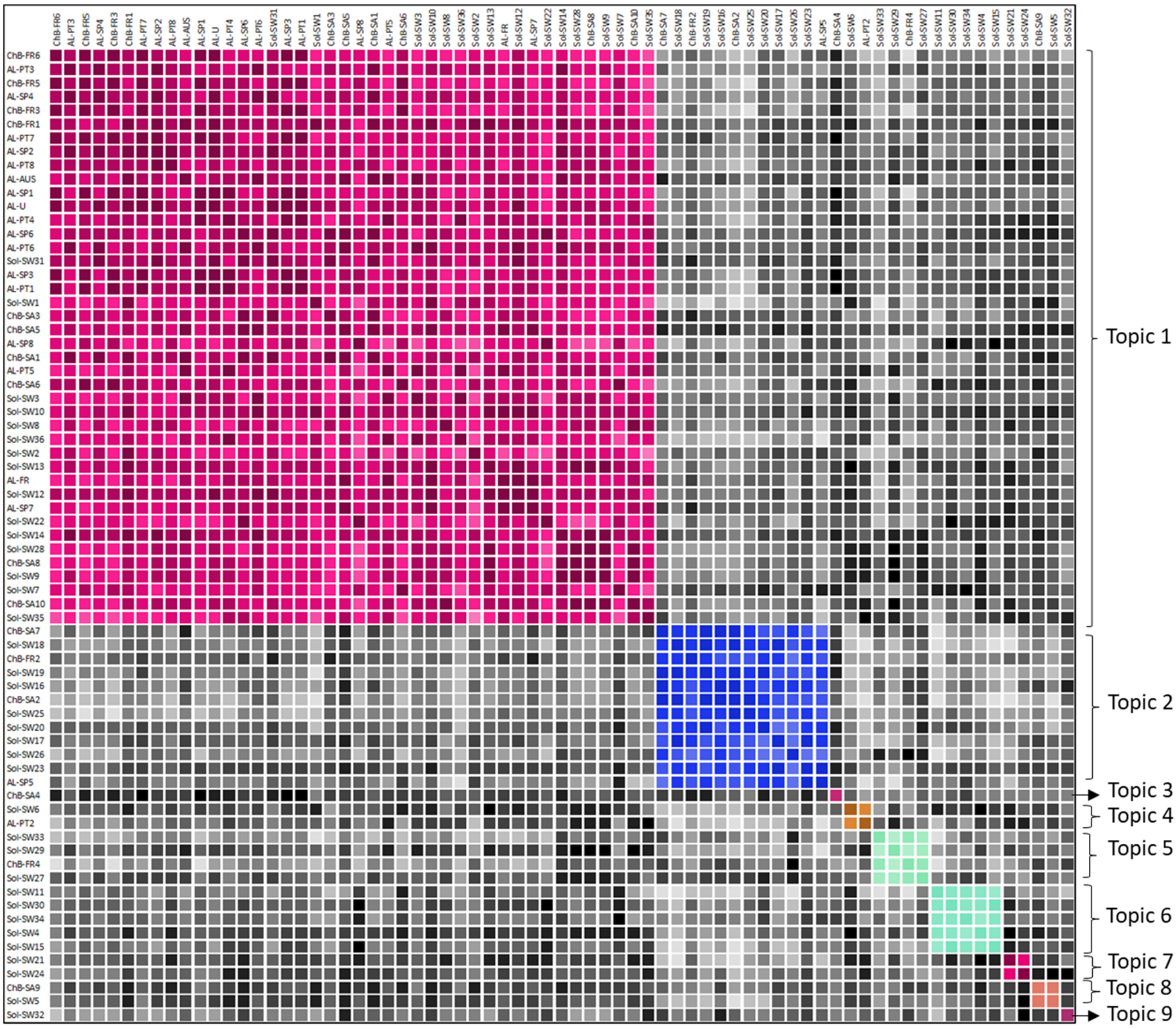

Figure 5. Wine-to-wine correlation matrix obtained from Latent Semantic Analysis (LSA). Wines are color-coded according to grape variety and country of origin.

Topic 2 included eight Solaris wines, three Chenin Blanc wines, and one Albariño wine, representing 67%, 25%, and 8% of the wines, respectively. All six of the Solaris wines previously associated with Topic 2 for the Solaris dataset are present here in the three-cultivar dataset. Although Topic 5 contains only Solaris wines, these were previously part of various topics in the Solaris dataset; similarly, the terms used for this topic were previously associated with various topics. When analyzing the remaining topics, it was found that SOL-SW32 (W32) stands on its own. This wine was previously associated with SOL-SW16 (W16).

The relationships between LSA topics were further explored for the dataset of the three cultivars through AHC on the MDS dimensions (Figure 6A,B). We found that Topic 1 was the most unique and dissimilar to the rest of the topics, followed by Topic 2. The rest of the topics were clustered together, indicating relatively high similarity between them (cophenetic correlation coefficient of 0.83). This cluster was further divided into two sub-clusters, Sub-Cluster 1 (Topics 3, 4, and 6) and Sub-Cluster 2 (Topics 4, 7, 8, and 9).

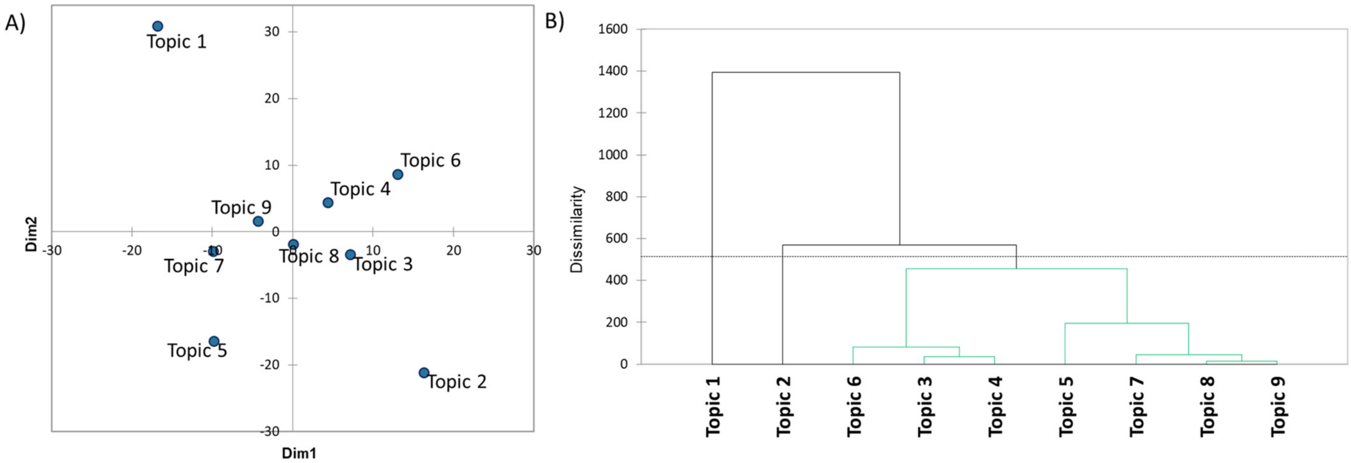

Figure 6. (**A**) Multidimensional Scaling (MDS) and (**B**) Agglomerative Hierarchical Clustering (AHC) of LSA topics for three-cultivar dataset.

3.2.2. Multiple Correspondence Analysis of Three Cultivars

The results from the MCA show that 70% of the explained variance was achieved over eleven dimensions, with 44% over the first two dimensions, as shown in Figure 7A. Figure 7A illustrates the distribution of all the wines (MCA scores plot) based on the presence (attribute-1)/absence (attribute-0) of the aroma descriptors. The projection of the two scores shows a more compact clustering of Albariño wines along the negative F1 axis, whereas Solaris and Chenin Blanc wines were more broadly distributed (Figure 7A). Of the two clusters in the AHC, the majority of the Albariño wines belonged to Cluster 2, with only one wine in Cluster 1 (cophenetic correlation coefficient of 0.485). Similarly to the score plot projections, the Solaris and Chenin Blanc wines varied throughout both clusters. This result is similar to the association of wines in the LSA, where most of the Albariño wines were associated with Topic 2. Thus, given the larger number of wines in Cluster 2, it is similar to Topic 1 of the LSA. Most of the Chenin Blanc wines previously associated with Topic 1 were present in Cluster 2. Cluster 2 also contained samples found in the other topics except for Topic 2, all 12 samples of which were present in Cluster 1.

The MCA loading plot (Figure 8) shows the top ten attributes with the highest contribution to F1, which are "apricot-1", "pear-0", "oak-1", "orange-1", "red apples-1", "butter-1", "Yellow plum-1", "rosehips-1", "pear-1", and "almond-1" in that order. They were mostly associated with Topic 2 in LSA, with the absence attribute "pear-0" being antithetical to "pear-1" and the term "pear", which were previously associated with Topic 1. The exceptions are "rosehips-1" and "almond-1", which were previously associated with Topic 9 associated with the Sol-SW32 Solaris wine; however, the MCA biplot shows that these attributes correlated with the vectors for Sol-SW33 and Sol-SW19.

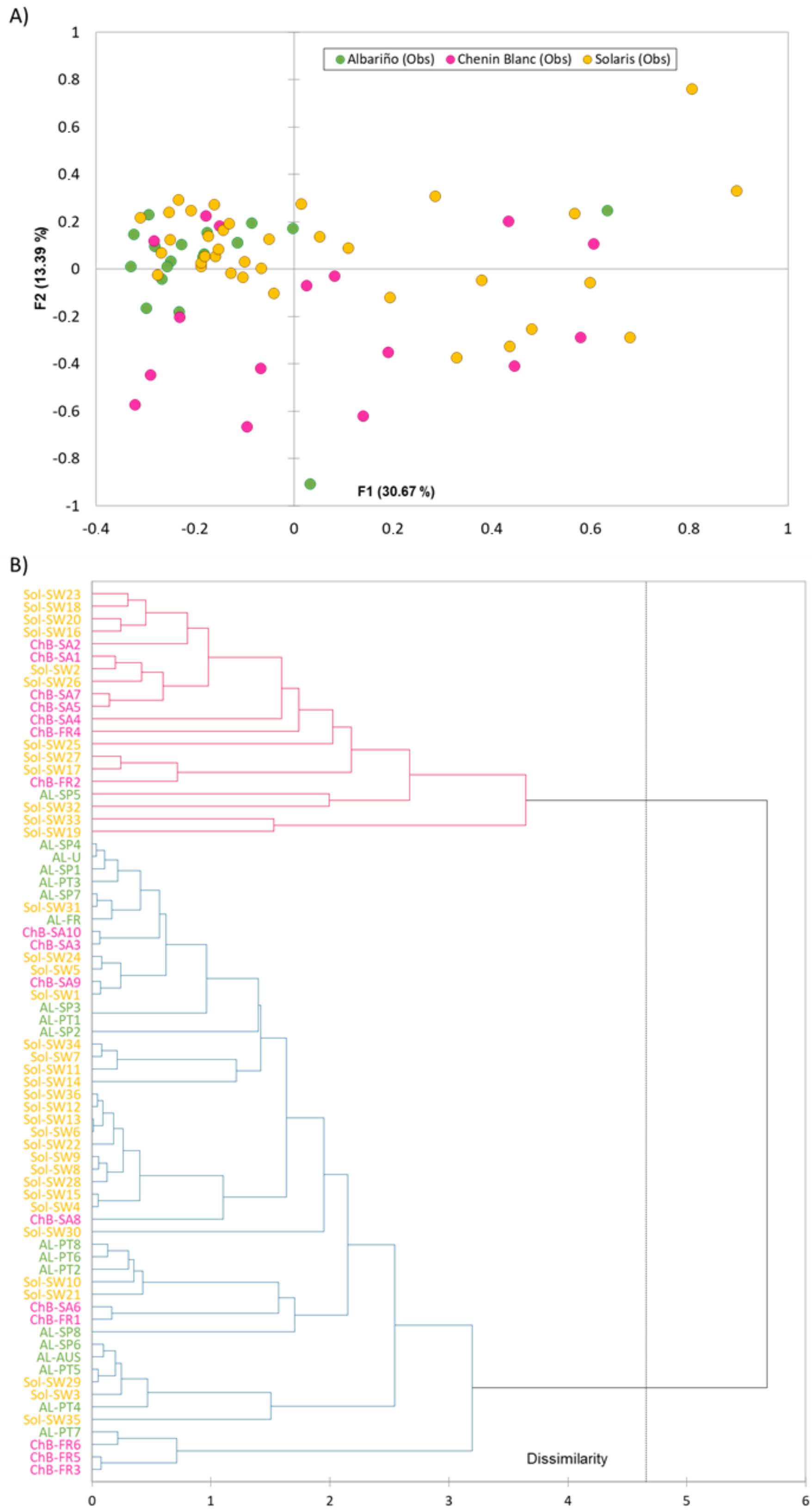

Figure 7. (**A**) Multiple Correspondence Analysis (MCA) biplot projecting wines based on their sensory descriptions and (**B**) corresponding Agglomerative Hierarchical Analysis (AHC). Solaris wines (Sol) are shown in yellow, Chenin Blanc wines (ChB) in pink, and Albariño wines (AL) in green.

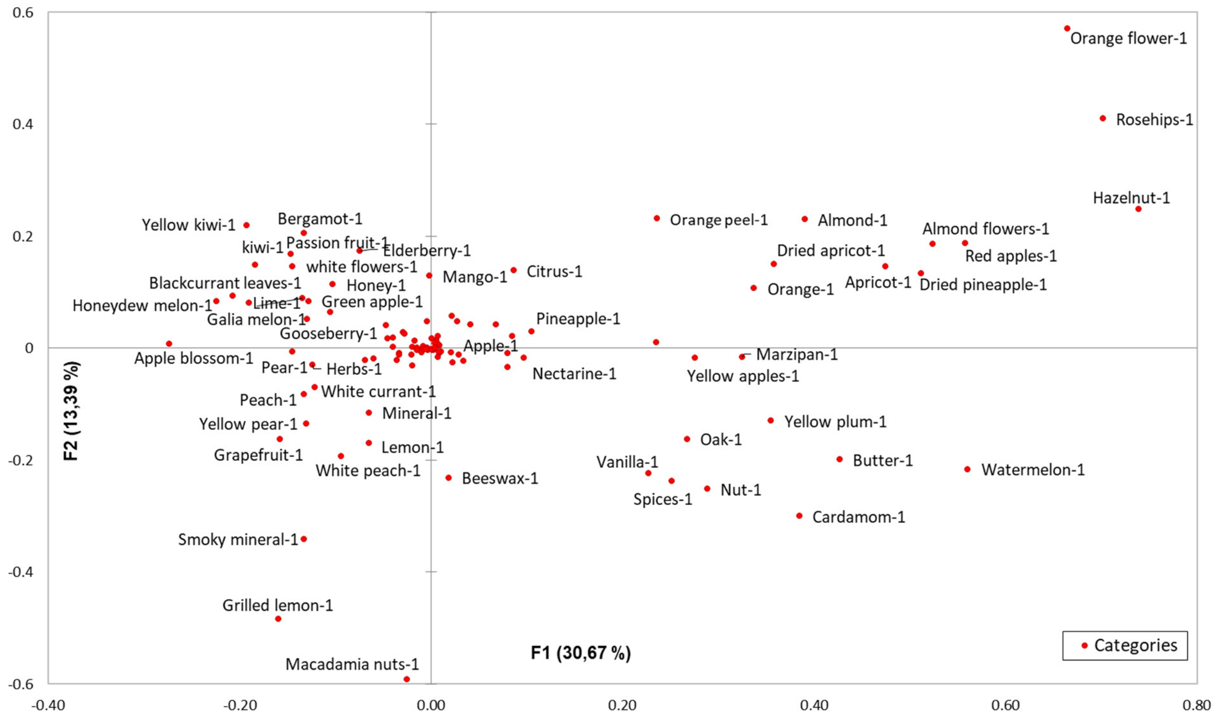

Figure 8. Loading plot of Multiple Correspondence Analysis (MCA) of three-cultivar dataset. For better visualization, only sensory attributes used to describe wines are represented.

4. Discussion

4.1. Suitability to Explore Swedish Solaris Wines

The present work explored LSA as an alternative tool for identifying potential sensory spaces from existing text data. LSA can be useful for understanding and potentially defining sensory spaces in addition to more established methods, such as CA and MCA. This comparison was carried out using cluster analysis (AHC). This comparison helped us understand the relevance and possible limitations of using LSA for sensory space investigations. Firstly, the Solaris dataset was analyzed, and with the addition of Chenin Blanc and Albariño wines, the sensory space was further comparatively investigated.

The results of the LSA on the Solaris dataset show that most wines were associated with two different topics (Topic 1 and Topic 2). Topic 1 mainly consisted of the terms related to white fruits, such as “pear” and “green apple”, but also “honey” and more generic descriptors such as “citrus” and “herbs” (Table 1). A total of 16 out of 36 wines were associated with this topic. Six Solaris wines were associated with Topic 2, which consisted of a larger number of terms such as “oak”, “apricot”, “butter”, and “yellow apple”. The rest of the Swedish Solaris wines (n = 13) were associated with LSA topics which were not found to be as representative based on topic weight (number of wines and/or terms) and percentage variability (Tables 1 and S1). A comparison of the topics using a cluster analysis (Figure 3) showed that Topics 1 and 3 were clustered together, while the rest of the topics were clustered together. Topics 1 and 3 have high dissimilarity, indicating that although they were clustered together, they could constitute distinguishable categories of Swedish Solaris wines. Therefore, this potentially indicates that Solaris wines could be distinguished into three possible styles based on Topics 1, 2, and 3.

The results of the MCA on the Solaris dataset show very similar clustering patterns. Two major clusters were found, with one compact cluster (Cluster 1; Figure 4B) indicating high similarity between the samples and another broad cluster (Cluster 2) being found due to high sample variation. Cluster 1 contained samples from LSA Topic 1, which were correlated with similar terms as those in the LSA (Figure 4A and Table 1). Likewise, Cluster 2 consisted of samples associated with LSA Topic 2 and correlated with similar terms used for this topic. Therefore, the LSA had similar cluster dynamics to MCA, with the additional dissemination of clustering of attributes/terms according to semantic topics. Although the MCA has more advantages over LSA for this dataset, the LSA improved the descriptive element in this study.

Nonetheless, since LSA prioritizes term-to-term associations over document (wine)-to-term/topic associations, the associations of wines with each topic were unreliable and unrepresentative of the original dataset. This limitation may be due to the low density in the dataset (the low numbers of wines and terms used for the descriptions). Another challenge faced when using LSA was finding the optimal number of documents and terms per topic to fit the data. Given the low data density, terms were often co-expressed in these wines, and by way of frequency, the most cited terms were more frequently associated. This was due to the low number of descriptors used for each wine and, overall, the low number of descriptors present. Hence, in both FoC and the percentage citation (Table S2), generally accepted optimization cut-offs for preprocessing could not be imposed since every available descriptor was meaningful.

The sample size also limited the possibility of exploring the data following a fuzzy clustering criterion. Sung et al. (2008) [52] already showed improved accuracy by using a fuzzy approach in LSA as it improved accuracy in text identification. However, the current sample size was a limitation to exploring a fuzzy approach. Nonetheless, the dataset is representative of a young Swedish Solaris wine industry with all commercially available Swedish Solaris wines on Systembolaget.

4.2. Suitability to Explore Potential Cultivar Typicality

From a Natural Language Processing (NLP) perspective, if we cannot find thematic/semantic similarities between Swedish Solaris wines using LSA, we may typify them by distinguishing them from other cultivars. This is why the second part of this study followed the same strategy to compare Swedish Solaris wines with monovarietal Chenin Blanc and Albariño wines. The LSA was performed over nine topics (cf., seven topics in the Solaris dataset) to allow for the incorporation of at least 70 percent of the variation for the larger three-cultivar dataset. The LSA performed on the three-cultivar dataset showed very similar results to the Solaris dataset in that two major topics were found, i.e., Topic 1 and Topic 2. Similarly, Topic 1 was described with the highest cited terms (i.e., "pear" and "herbs"), while Topic 2 had the second highest cited terms associated with it.

Again, Topic 1 contained the majority of all samples, while Topic 2 had the second greatest number of samples. The association of Solaris wines with topics was similar to that in the Solaris-only dataset. This is reasonable since the three-cultivar dataset had predominantly Swedish Solaris wines (i.e., 36 out of 71 wines). Topic 1 had the majority of Solaris wines that were associated with the Solaris-only dataset. Topic 1 also contained the majority of Albariño wines. This suggests that Topic 1 may be a generic topic based on highly cited terms and is not a good distinguishing topic among samples.

Conversely, Topic 2 had most of the terms associated with Topic 2 of the Solaris dataset, as well as all of the Solaris wines that were associated with it. Comparatively, it seems to better distinguish Solaris wines from the other cultivars since the majority of wines associated with this topic were Solaris wines. A cluster analysis on the topics showed that Topic 1 and Topic 2 were very dissimilar from the rest of the topics which clustered together (Figure 6). This may indicate that although the Solaris dataset indicated that there were potentially three categories of Swedish Solaris wines, in comparison with two similar international white wines (i.e., Chenin Blanc and Albariño), they could only be distinguished from the rest based on Topic 2. MCA supported the LSA results, and AHC of the scores showed that Cluster 1 contained all of the Topic 2 wines (Figure 7B). Furthermore, the MCA biplot showed that the sample vectors correlated with attributes similar to those associated with LSA Topic 2. However, since the majority of samples were Swedish Solaris wines, perhaps creating a more balanced dataset (between cultivars) could create a more fair and representative distribution of samples across the different topics and clusters.

Some studies have proposed that class imbalances can influence cluster and classification results [45]. In the case of LSA, clustering is performed on the raw data, whereas MCA can be followed by clustering on the resultant correlation matrix. Although both LSA and MCA work on sparse data, LSA emphasizes the relationships between attributes, while MCA balances the relationships of both the samples and the attributes, increasing the informational output and reducing the imbalance issue. Thus, although LSA allows for valuable semantic clustering of descriptors, the numerous limitations found in this study highlight the cautionary measures that need to be taken when interpreting the results. Additionally, MCA has more advantages over LSA, indicating that LSA may only be used in the form of an annex technique rather than as the primary analysis method.

From an overall sensory perspective, the results show that the majority of Solaris wines were frequently characterized by the presence of white fruit aromas ("pear" and "green apple") in combination with citrus and greener notes such as "herbs". However, in comparison with Chenin Blanc and Albariño wines, these were shown to be generic descriptors used for all three cultivars, and they could not distinguish Solaris wines from the rest. However, Albariño wines were better described by these generic descriptors compared to the other cultivars. This may indicate that, although these terms are very common among these cultivars, Albariño wines are more typically described with them [53]. Hence, these terms could be used for the classification of Albariño wines, but Chenin Blanc and Solaris wines had more diverse sensory descriptor profiles. For cultivars with different styles (e.g., Chenin Blanc styles such as *fruity*, green, and *mature*), variation in the sensory space is expected, hence explaining the Chenin Blanc results observed in this study. Comparatively, this may be the case for Solaris, which was the primary cultivar of investigation. Solaris wines in this study consistently showed an association/correlation with descriptors (i.e., "oak", "buttery", "apricot", and "red apples"), which may indicate a distinct style of Solaris wines. However, the concept of typicality of Swedish Solaris wines (descriptors distinguishing all Solaris wines from the other cultivars) could not be defined.

4.3. General Discussion and Conclusions

The present study provides insights into the sensory space of Solaris wines, a promising international white grape cultivar. Furthermore, the study investigated the differences between the Swedish Solaris wines currently available on the Swedish market. The results also show the potential use and limitations of LSA as a tool to investigate sensory spaces from existing text data compared to a more conventional multivariate analysis. Similar cluster dynamics were found between LSA and MCA. However, LSA provides additional sensory insights thanks to the clustering of attributes/terms according to semantic topics.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/beverages10040120/s1, Table S1: Distribution and variability among topics for LSA; Table S2: Descriptor variation in three-cultivar dataset.

Author Contributions: Conceptualization: G.G.-B., M.M. and A.B.; Data Curation: M.M. and G.G.-B.; Investigation: G.G.-B., M.M. and A.B.; Methodology: G.G.-B., M.M. and A.B.; Software: M.M. and G.G.-B.; Validation: G.G.-B., M.M. and A.B.; Visualization: G.G.-B., M.M. and A.B.; Writing—Original Draft: G.G.-B., M.M. and A.B.; Writing—Review and Editing: G.G.-B., M.M. and A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Commins, T.; Asavasanti, S.; Deloire, A. What Is Tropical Wine and What Defines It? Thailand as a Case Study. *Asian J. Food Agro-Ind.* **2012**, *5*, 79–95.
2. de Souza Nascimento, A.M.; de Souza, J.F.; Costa, A.O.; de Freitas Santos, S.; Silva, G.G.; Pereira, G.E. Caracterization of Phenolic Composition of Altitude Tropical Wines in the Brazilian Northeast. *BIO Web Conf.* **2016**, *7*, 02014. [CrossRef]

3. Pereira, G.E.; Araújo, A.J.B.; De Santos, J.O.; Oliveira, V.S.; Nascimento, R.L.; Quintino, C.; Vanderlinde, R.; Lima, L.L.A. Chemical and Aromatic Characteristics of Brazilian Tropical Wines. In Proceedings of the Acta Horticulturae; International Society for Horticultural Science, 31 October 2011; Volume 910, pp. 135–140.
4. Nordmark, L.; Lindén, J.; Skjöldebrand, C.; Hansson, H. The Nordic Light Terroir. *Acta Hortic.* **2016**, *1115*, 189–193. [CrossRef]
5. Liu, J.; Toldam-Andersen, T.B.; Petersen, M.A.; Zhang, S.; Arneborg, N.; Bredie, W.L.P. Instrumental and Sensory Characterisation of Solaris White Wines in Denmark. *Food Chem.* **2015**, *166*, 133–142. [CrossRef]
6. Bjenning, L. Mapping Sweden as a Wine Country. Master's Thesis, Kristianstad University, Kristianstad, Sweden, 2024.
7. Pedneault, K.; Provost, C. Fungus Resistant Grape Varieties as a Suitable Alternative for Organic Wine Production: Benefits, Limits, and Challenges. *Sci. Hortic.* **2016**, *208*, 57–77. [CrossRef]
8. Ruehl, E.; Schmid, J.; Eibach, R.; Töpfer, R. *Grapevine Breeding Programmes in Germany*; Elsevier Ltd.: Amsterdam, The Netherlands, 2015; ISBN 9781782420804.
9. Ehrhardt, C.; Arapitsas, P.; Stefanini, M. Analysis of the Phenolic Composition of Fungus- Resistant Grape Varieties Cultivated in Italy and Germany Using UHPLC-MS/MS†. *J. Mass Spectrom.* **2014**, *49*, 860–869. [CrossRef]
10. Vivino Vivino. Available online: https://www.vivino.com/SE/sv/ (accessed on 23 June 2021).
11. Garrido-Bañuelos, G.; Ballester, J.; Buica, A.; Mihnea, M. Exploring the Typicality, Sensory Space, and Chemical Composition of Swedish Solariswines. *Foods* **2020**, *9*, 1107. [CrossRef] [PubMed]
12. Ruiz-Capillas, C.; Herrero, A.M. Sensory Analysis and Consumer Research in New Product Development. *Foods* **2021**, *10*, 582. [CrossRef]
13. Ballester, J.; Patris, B.; Symoneaux, R.; Valentin, D. Conceptual vs. Perceptual Wine Spaces: Does Expertise Matter? *Food Qual. Prefer.* **2008**, *19*, 267–276. [CrossRef]
14. Barbe, J.C.; Garbay, J.; Tempère, S. The Sensory Space of Wines: From Concept to Evaluation and Description. A Review. *Foods* **2021**, *10*, 1424. [CrossRef]
15. Mafata, M.; Brand, J.; Panzeri, V.; Buica, A. Investigating the Concept of South African Old Vine Chenin Blan. *S. Afr. J. Enol. Vitic.* **2020**, *14*, 168–182. [CrossRef]
16. Rodrigues, H.; Gomez-Corona, C. "It Looks like a Garden": Wine Descriptions in Social Media—The Case of the Sensory Descriptor "Floral" in 13 English-Speaking Countries. In Proceedings of the EuroSense 2020, Berlin, Germany, 13–16 September 2020.
17. Ickes, C.M.; Lee, S.Y.; Cadwallader, K.R. Novel Creation of a Rum Flavor Lexicon Through the Use of Web-Based Material. *J. Food Sci.* **2017**, *82*, 1216–1223. [CrossRef] [PubMed]
18. Hamilton, L.M.; Lahne, J. Fast and Automated Sensory Analysis: Using Natural Language Processing for Descriptive Lexicon Development. *Food Qual. Prefer.* **2020**, *83*, 103926. [CrossRef]
19. Da Silva, T.M.; Marinoni, D.T.; Peano, C.; Giuggioli, N.R. A New Sensory Approach Combined with a Text-Mining Tool to Create a Sensory Lexicon and Profile of Monovarietal Apple Juices. *Foods* **2019**, *8*, 608. [CrossRef]
20. Garrido-Bañuelos, G.; de Barros Alves, H.; Mihnea, M. Mapping the Sensory Fingerprint of Swedish Beer Market through Text and Data Mining and Multivariate Strategies. *Beverages* **2021**, *7*, 74. [CrossRef]
21. Valente, C.C.; Bauer, F.F.; Venter, F.; Watson, B.; Nieuwoudt, H.H. Modelling the Sensory Space of Varietal Wines: Mining of Large, Unstructured Text Data and Visualisation of Style Patterns. *Sci. Rep.* **2018**, *8*, 4987. [CrossRef] [PubMed]
22. Platter's Wine Guide. Available online: https://www.wineonaplatter.com/ (accessed on 23 June 2021).
23. Croijmans, I.; Hendrickx, I.; Lefever, E.; Majid, A.; van Den Bosch, A. Uncovering the Language of Wine Experts. *Nat. Lang. Eng.* **2019**, *26*, 511–530. [CrossRef]
24. Hendrickx, I.; Lefever, E.; Croijmans, I.; Majid, A.; Van Den Bosch, A. Very Quaffable and Great Fun: Applying NLP to Wine Reviews. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), Berlin, Germany, 7–12 August 2016.
25. Chen, B.; Velchev, V.; Palmer, J.; Atkison, T. Wineinformatics: A Quantitative Analysis of Wine Reviewers. *Fermentation* **2018**, *4*, 82. [CrossRef]
26. Brand, J.; Panzeri, V.; Buica, A. Wine Quality Drivers: A Case Study on South African Chenin Blanc and Pinotage Wines. *Foods* **2020**, *9*, 805. [CrossRef]
27. Varela, P.; Ares, G. Sensory Profiling, the Blurred Line between Sensory and Consumer Science. A Review of Novel Methods for Product Characterization. *Food Res. Int.* **2012**, *48*, 893–908. [CrossRef]
28. Mafata, M.; Brand, J.; Medvedovici, A.; Buica, A. Chemometric and Sensometric Techniques in Enological Data Analysis. *Crit. Rev. Food Sci. Nutr.* **2022**, *63*, 10995–11009. [CrossRef] [PubMed]
29. Vera, L.; Aceña, L.; Guasch, J.; Boqué, R.; Mestres, M.; Busto, O. Discrimination and Sensory Description of Beers through Data Fusion. *Talanta* **2011**, *87*, 136–142. [CrossRef] [PubMed]
30. Biancolillo, A.; Boqué, R.; Cocchi, M.; Marini, F. Data Fusion Strategies in Food Analysis. *Data Handl. Sci. Technol.* **2019**, *31*, 271–310. [CrossRef]
31. Garrido-Bañuelos, G.; Buica, A. Is There a Link Between Coffee Aroma and the Level of Furanmethanethiol (FMT) in Pinotage Wines ? *S. Afr. J. Enol. Vitic.* **2020**, *41*, 245–250. [CrossRef]

32. Dong, J.J.; Li, Q.L.; Yin, H.; Zhong, C.; Hao, J.G.; Yang, P.F.; Tian, Y.H.; Jia, S.R. Predictive Analysis of Beer Quality by Correlating Sensory Evaluation with Higher Alcohol and Ester Production Using Multivariate Statistics Methods. *Food Chem.* **2014**, *161*, 376–382. [CrossRef]
33. Esti, M.; Airola, R.L.G.; Moneta, E.; Paperaio, M.; Sinesio, F. Qualitative Data Analysis for an Exploratory Sensory Study of Grechetto Wine. *Anal. Chim. Acta* **2010**, *660*, 63–67. [CrossRef]
34. McEwan, J.A.; Schlich, P. Correspondence Analysis in Sensory Evaluation. *Food Qual. Prefer.* **1991**, *3*, 23–36. [CrossRef]
35. Meyners, M.; Castura, J.C.; Carr, B.T. Existing and New Approaches for the Analysis of CATA Data. *Food Qual. Prefer.* **2013**, *30*, 309–319. [CrossRef]
36. Abdi, H.; Valentin, D. Multiple Correspondence Analysis. In *Encyclopedia of Measurement and Statistics*; SAGE Publications: Thousand Oaks, CA, USA, 2007.
37. Fuentes, S.; Torrico, D.D.; Tongson, E.; Viejo, C.G. Machine Learning Modeling of Wine Sensory Profiles and Color of Vertical Vintages of Pinot Noir Based on Chemical Fingerprinting, Weather and Management Data. *Sensors* **2020**, *20*, 3618. [CrossRef]
38. Aleixandre-Tudó, J.L.; Álvarez, I.; Lizama, V.; Nieuwoudt, H.H.; García, M.J.; Aleixandre, J.L.; du Toit, W. Modelling Phenolic and Volatile Composition to Characterize the Effects of Pre-Fermentative Cold Soaking in Tempranillo Wines. *LWT—Food Sci. Technol.* **2016**, *66*, 193–200. [CrossRef]
39. Bhardwaj, P.; Tiwari, P.; Olejar, K.; Parr, W.; Kulasiri, D. A Machine Learning Application in Wine Quality Prediction. *Mach. Learn. Appl.* **2022**, *8*, 100261. [CrossRef]
40. Villière, A.; Fillonneau, C.; Prost, C.; Guillet, F. SketchOscent: Towards a Knowledge-Based Model and Interactive Visualisation of the Odour Space. In Proceedings of the 16thWeurman Flavour Research Symposium, 21–24 September 2021.
41. Liu, C.; Shang, L.; Hayashi, K. Co-Occurrence-Based Clustering of Odor Descriptors for Predicting Structure-Odor Relationship. In Proceedings of the ISOEN 2019—18th International Symposium on Olfaction and Electronic Nose, Proceedings, Fukuoka, Japan, 26–29 May 2019.
42. Evangelopoulos, N.E. Latent Semantic Analysis. *Wiley Interdiscip. Rev. Cogn. Sci.* **2013**, *4*, 683–692. [CrossRef] [PubMed]
43. Ahmad, S.N.; Laroche, M. How Do Expressed Emotions Affect the Helpfulness of a Product Review? Evidence from Reviews Using Latent Semantic Analysis. *Int. J. Electron. Commer.* **2016**, *20*, 76–111. [CrossRef]
44. Landauer, T.K.; Foltz, P.W.; Laham, P.W. Introduction to Latent Semantic Analysis. *Discourse Process.* **1998**, *25*, 259–284. [CrossRef]
45. Velldal, E. *Modeling Word Senses With Fuzzy Clustering*; University of Oslo: Oslo, Norway, 2003.
46. Systembolaget. Available online: https://www.systembolaget.se/ (accessed on 23 June 2021).
47. Mafata, M.; Williams, C.; Kruger, M.; Brand, J.; Watson, B.; Buica, A. Unsupervised Data Fusion and Interpretation through Cluster Analysis on Biplot Projections: Craft Beer and Gin Case Studies. *Eur. Food Res. Technol.* **2023**, *249*, 1073–1087. [CrossRef]
48. Schreiber, J.B. Issues and Recommendations for Exploratory Factor Analysis and Principal Component Analysis. *Res. Social Adm. Pharm.* **2021**, *17*, 1004–1011. [CrossRef]
49. Howard, M.C. A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? *Int. J. Hum. Comput. Interact.* **2016**, *32*, 51–62. [CrossRef]
50. Mafata, M.; Brand, J.; Buica, A. Data Fusion Using Multiple Factor Analysis Coupled with Non-Linear Pattern Recognition (Fuzzy k-Means): Application to Chenin Blanc. *OENO One* **2022**, *56*, 413–425. [CrossRef]
51. McKillup, S. *Statistics Explained: An Introductory Guide for Life Scientists*; Cambridge University Press: Cambridge, UK, 2011. [CrossRef]
52. Sun, J.; Zhang, Q.; Yuan, Z.; Huang, W.; Yan, X.; Dong, J. Calculation of Latent Semantic Weight Based on Fuzzy Membership. In Proceedings of the Advances in Neural Networks-ISNN 2008: 5th International Symposium on Neural Networks, ISNN 2008, Beijing, China, 24–28 September 2008; Proceedings, Part II 5. Springer: Berlin/Heidelberg, Germany, 2008; pp. 91–99. [CrossRef]
53. Vilanova, M.; Zamuz, S.; Tardáguila, J.; Masa, A. Descriptive Analysis of Wines from Vitis Vinifera Cv. Albariño. *J. Sci. Food Agric.* **2008**, *88*, 819–823. [CrossRef]