

# Les vins de Loire du cépage Chenin ont-ils un goût identifiable ?

Jérôme Pagès et François Husson

Même la dégustation des vins gagne à être soumise à l'analyse statistique ! Lors d'une évaluation sensorielle de dix vins par douze professionnels, on a demandé à chaque dégustateur de fournir, pour chaque vin, quelques mots le caractérisant. L'analyse des correspondances permet de synthétiser les associations entre mots et vins et ouvre la voie aux interprétations de ces avis d'experts.

Cette étude s'inscrit dans une recherche sur la typicité des vins provenant du cépage Chenin en Val de Loire<sup>1</sup>. L'angle adopté ici est de situer la variabilité des chenins au sein de la variabilité des vins blancs du Val de Loire. Pour cela, il a été décidé de confronter des chenins à des vins issus de l'autre grand cépage blanc du Val de Loire, le Sauvignon. Cette confrontation repose sur la caractérisation de ces vins par un ensemble de dégustateurs. L'hypothèse est que la diversité des caractérisations obtenues correspond à la différence de cépage.

Concrètement, dix vins blancs secs de Touraine, cinq Touraine AOC, issus du cépage Sauvignon, et cinq Vouvray AOC, issus du cépage Chenin (cf. Tableau 1) ont été choisis par J.P. Gouvazé (Interloire, Tours), la demande étant d'illustrer la diversité, au sein de chaque cépage, des vins de Loire. Remarquons qu'il y a une confusion totale (au sens de la planification expérimentale) entre l'appellation et le cépage. Par la suite, pour simplifier, nous ne parlerons que du facteur cépage.



**Tableau 1. Les dix vins étudiés.**

| Numéro | Nom                  | Cépage    | Appellation | Remarque               |
|--------|----------------------|-----------|-------------|------------------------|
| 1      | Michaud              | Sauvignon | Touraine    |                        |
| 2      | Renaudie             | Sauvignon | Touraine    |                        |
| 3      | Trotignon            | Sauvignon | Touraine    |                        |
| 4      | Buisse Domaine       | Sauvignon | Touraine    |                        |
| 5      | Buisse Cristal       | Sauvignon | Touraine    |                        |
| 6      | Aubuissières Silex   | Chenin    | Vouvray     | 7g de sucres résiduels |
| 7      | Aubuissières Marigny | Chenin    | Vouvray     | élevage en fût         |
| 8      | Fontainerie Domaine  | Chenin    | Vouvray     |                        |
| 9      | Fontainerie Brûlés   | Chenin    | Vouvray     |                        |
| 10     | Fontainerie Coteaux  | Chenin    | Vouvray     | élevage en fût         |

Les données analysées ici proviennent d'une dégustation rassemblant douze professionnels auxquels il était demandé, pour chaque vin, de donner un (ou plusieurs) mot(s) qui caractérisai(ent) bien sensoriellement chaque vin. Comme il est d'usage, la dégustation a eu lieu « à l'aveugle », les vins étant présentés sans aucune indication. Toutefois, cette dégustation ayant eu lieu lors d'un Salon des vins de Loire, les dégustateurs étaient fondés à penser qu'il s'agissait de vins de Loire même si cela ne leur a pas été dit ; mais la diversité des vins de Loire, du point de vue des terroirs, des cépages et des pratiques viti-vinicoles, autorise à penser que ce niveau d'information très général sur l'ensemble des vins est sans effet sur les caractérisations relatives des vins.

## Les vins et les mots

Il s'agit donc en quelque sorte d'un questionnaire comportant dix questions ouvertes (une par vin). Ces données sont rassemblées dans un tableau dont les lignes sont les vins, dont les colonnes sont les mots et dont le terme général  $x_{ij}$  est le nombre de fois que le mot  $j$  a été associé au vin  $i$  (cf. Tableau 2).

**Tableau 2.**

Tableau des données : nombre de fois que chaque mot a été utilisé pour chaque vin (30 mots).

|                  | Fruité | Sucré, doux, liquoreux | Bois, boisé | Léger, souple | Acidité | Agrume | Jaune doré | Vif | Arômes fruités | Fin, discret | Amertume | Floral | Gras, manque de fraîcheur | Jaune pâle, clair | Frais | Long en bouche, persistant | Floral, fleurs blanches | Sec | Intense, ample | Miel | Nez fermé, faible | Ouvert, expressif | Bonne expression aromatique | Goût étranger (cire, pneu) | Goût vert | lode | Peu d'acidité | Peu de caractère, d'expression | Sauvignon | Odeur soutenue | Somme |
|------------------|--------|------------------------|-------------|---------------|---------|--------|------------|-----|----------------|--------------|----------|--------|---------------------------|-------------------|-------|----------------------------|-------------------------|-----|----------------|------|-------------------|-------------------|-----------------------------|----------------------------|-----------|------|---------------|--------------------------------|-----------|----------------|-------|
| 1S-Michaud       | 1      | 0                      | 1           | 1             | 1       | 2      | 2          | 2   | 2              | 0            | 1        | 0      | 0                         | 1                 | 1     | 1                          | 2                       | 0   | 1              | 0    | 0                 | 2                 | 1                           | 0                          | 2         | 1    | 1             | 0                              | 1         | 1              | 28    |
| 2S-Renaudie      | 5      | 1                      | 0           | 0             | 0       | 3      | 0          | 3   | 1              | 2            | 1        | 1      | 0                         | 2                 | 2     | 1                          | 1                       | 0   | 0              | 1    | 0                 | 0                 | 1                           | 0                          | 0         | 1    | 0             | 0                              | 1         | 0              | 27    |
| 3S-Trotignon     | 5      | 1                      | 0           | 2             | 1       | 1      | 0          | 0   | 2              | 1            | 0        | 2      | 0                         | 2                 | 2     | 1                          | 1                       | 0   | 0              | 0    | 0                 | 1                 | 1                           | 0                          | 2         | 0    | 0             | 0                              | 1         | 0              | 26    |
| 4S-Buisse        | 3      | 0                      | 0           | 2             | 2       | 1      | 1          | 1   | 1              | 4            | 0        | 0      | 0                         | 0                 | 2     | 0                          | 0                       | 3   | 0              | 0    | 0                 | 0                 | 1                           | 0                          | 0         | 1    | 1             | 0                              | 0         | 0              | 23    |
| 5S-BuisseCristal | 4      | 0                      | 2           | 1             | 1       | 1      | 0          | 3   | 0              | 0            | 0        | 2      | 0                         | 1                 | 0     | 0                          | 1                       | 1   | 0              | 0    | 0                 | 0                 | 0                           | 0                          | 0         | 1    | 2             | 0                              | 0         | 2              | 22    |
| 6C-AubSilex      | 0      | 11                     | 0           | 2             | 0       | 0      | 1          | 1   | 1              | 0            | 0        | 0      | 2                         | 2                 | 0     | 0                          | 0                       | 0   | 1              | 1    | 3                 | 0                 | 0                           | 0                          | 0         | 0    | 0             | 1                              | 0         | 0              | 26    |
| 7C-AubMarigny    | 1      | 1                      | 7           | 0             | 2       | 0      | 2          | 1   | 0              | 0            | 0        | 0      | 2                         | 0                 | 0     | 2                          | 0                       | 0   | 1              | 1    | 0                 | 0                 | 0                           | 3                          | 0         | 0    | 0             | 2                              | 0         | 1              | 26    |
| 8C-FontDomaine   | 4      | 2                      | 0           | 0             | 1       | 3      | 1          | 0   | 1              | 1            | 1        | 1      | 1                         | 0                 | 0     | 0                          | 0                       | 1   | 0              | 1    | 2                 | 1                 | 0                           | 0                          | 0         | 0    | 0             | 0                              | 0         | 0              | 21    |
| 9C-FontBrûlés    | 3      | 1                      | 1           | 4             | 2       | 0      | 2          | 0   | 1              | 1            | 2        | 1      | 2                         | 0                 | 0     | 1                          | 0                       | 1   | 1              | 1    | 0                 | 1                 | 0                           | 0                          | 0         | 0    | 0             | 1                              | 0         | 0              | 26    |
| 10C-FontCoteaux  | 1      | 1                      | 5           | 0             | 1       | 0      | 2          | 0   | 0              | 0            | 3        | 1      | 1                         | 0                 | 0     | 1                          | 1                       | 0   | 1              | 0    | 0                 | 0                 | 0                           | 1                          | 0         | 0    | 0             | 0                              | 1         | 0              | 20    |
| Somme            | 27     | 18                     | 16          | 12            | 11      | 11     | 11         | 11  | 9              | 9            | 8        | 8      | 8                         | 8                 | 7     | 7                          | 6                       | 6   | 5              | 5    | 5                 | 5                 | 4                           | 4                          | 4         | 4    | 4             | 4                              | 4         | 4              | 245   |

L'objectif de l'analyse d'un tel tableau est d'étudier la liaison qu'il y a entre les mots d'une part et les vins d'autre part. S'il n'y a pas de liaison, alors les mots sont utilisés indifféremment selon les vins : concrètement, dans ce cas, les lignes du tableau sont à peu près identiques (le nombre de mots par vin est presque constant, cf. plus loin) et les colonnes sont proportionnelles. Dans le cas contraire, la présence d'une liaison signifie qu'il y a des associations privilégiées entre certains mots et certains vins. Ce sont ces associations que nous cherchons à mettre en évidence de façon synthétique.

Ce type de tableau est fréquemment constitué et une méthode adaptée à son analyse est l'Analyse des Correspondances (historiquement, la première analyse des correspondances publiée, dans la thèse de Brigitte Escofier en 1965, portait sur un tableau analogue). Mais d'habitude on dispose d'effectifs beaucoup plus importants. Nous sommes ici dans des conditions limites du fait d'un nombre total d'occurrences ( $n=245$ ) très faible. Néanmoins, l'analyse est possible du fait du vocabulaire plutôt stéréotypé des professionnels du vin, ce qui conduit à un nombre total de mots pas trop élevé et donc un nombre « suffisant » de mots présentant un effectif pas trop faible. En outre, en préalable à l'analyse, quelques mots « voisins » ont été regroupés

(par exemple sucré, doux et liquoreux qui renvoient à la même perception, celle de la saveur sucrée). Dans ce texte, dans un souci de simplification, nous gardons le terme « mot » pour les colonnes du tableau 2 même lorsqu'elles représentent des groupes de mots figurant tels quels dans les questionnaires (e.g. manque de fraîcheur) ou issus d'un regroupement a posteriori (sucré, doux).

Dans ce type d'analyse, on élimine les mots les moins utilisés. Compte tenu des faibles effectifs, la limite a été fixée à 4, seuil en dessous duquel les mots ne sont pas pris en compte. La détermination de ce seuil revêt toujours un caractère empirique : avec ces données, mettre le seuil à 5 ne change pas fondamentalement la représentation des vins mais prive de mots importants (par exemple « Sauvignon ») et le seuil de 3 conduit à des graphiques très chargés, qui plus est de mots dont la position sur le plan est fragile.

L'objectif principal de cette analyse est de fournir une image synthétique de la diversité de ces vins. Mais cette étude présente aussi un caractère méthodologique : est-il possible d'obtenir avec un protocole très simplifié (les dégustateurs ne sont pas entraînés ensemble ; ils utilisent leur propre vocabulaire) une image intéressante de la diversité des vins ?

### « Fruité » est le mot le plus utilisé

L'examen des marges (sommes des nombres d'une même ligne ou d'une même colonne) est important à la fois du point de vue de leur interprétation directe (quels sont les mots les plus utilisés? Certains vins font-ils l'objet de commentaires plus nombreux que d'autres?) et de celui de leur influence en analyse des correspondances (les mots les plus utilisés auront un poids plus important dans l'analyse).



Le mot le plus utilisé est « fruité » ce qui est en phase avec l'observation courante de commentaires de dégustation (pour s'en convaincre lire la contre-étiquette de n'importe quelle bouteille : il est difficile d'échapper au « merveilleux fruité » ...). Vient ensuite, l'ensemble {sucré, doux, liquoreux}. Rappelons que ces vins sont secs et une perception de sucré est de facto une caractéristique remarquable. Enfin, la perception de bois, associée à un élevage en fût, est bien identifiée par les professionnels, ce qui favorise l'apparition d'une fréquence de citation élevée (par opposition à une odeur facilement perçue mais non reconnue qui engendre des mots différents selon les dégustateurs). Sur le plan technique, du point de vue de l'analyse des correspondances, les mots auront, à profil égal, un poids d'autant plus important qu'ils ont été fréquemment cités, ce qui est souhaitable.

En revanche, les nombres de mots par vin semblent homogènes. Aucun vin ne semble attirer plus de commentaires que d'autres, ce qui est sans doute une conséquence (souhaitée) de la forme de la question posée (« Pour chaque vin, donner un ou quelques mots... »). Par acquis de conscience, on peut tester l'égalité du nombre de mots utilisés par vin grâce à un test d'ajustement des dix effectifs observés (dernière colonne du tableau 2) à une loi uniforme. La probabilité critique (0.97) confirme qu'il ne faut pas accorder d'attention aux différences entre les effectifs des mots par vin. Du point de vue de l'analyse des correspondances, on

peut considérer que les vins auront, à profil égal, à peu près la même influence dans l'analyse. Lorsque le nombre de mots par vin est différent, l'analyse accorde à un vin un poids d'autant plus important qu'il a fait l'objet de nombreux commentaires (son profil est mieux connu).

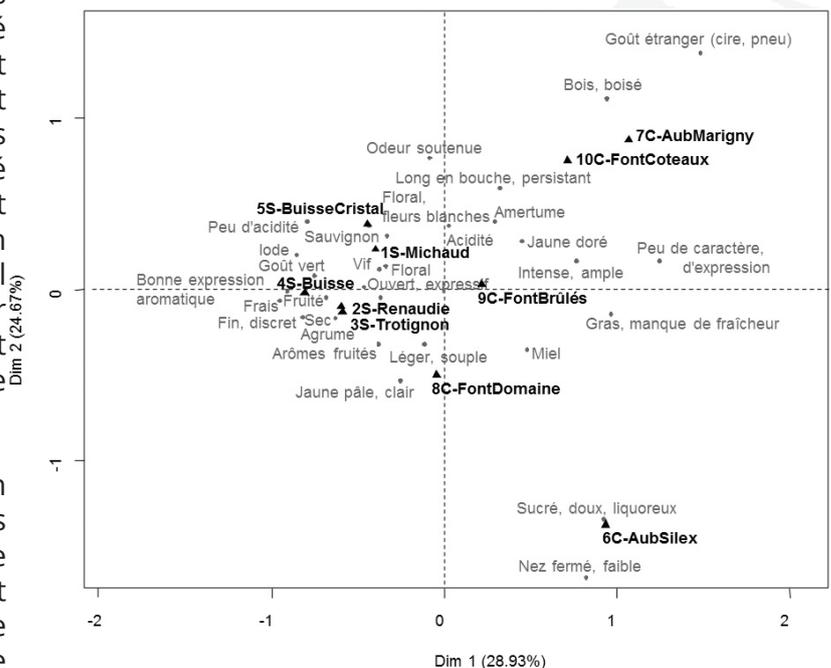
### Les mots choisis permettent-ils de caractériser les vins?

Sans entrer dans les détails techniques, il existe un test statistique (le test du Chi<sup>2</sup> d'indépendance) indiquant si la structure des données (c'est-à-dire ici l'association des mots et des vins) peut être considérée comme le produit du hasard ou si, au contraire, il existe des associations entre mots et vins suffisamment fortes pour mériter l'attention. Bien que nous ne soyons pas dans les conditions d'application du test du Chi<sup>2</sup> d'indépendance, nous donnons la décision du test à titre d'indication. Ici ce test est positif : le tableau mérite l'attention. Remarquons que, pour qui connaît la diversité entre les vins de Loire (surtout compte tenu du fait que ces vins ont été choisis pour illustrer cette diversité), la liaison entre les mots et les vins est attendue. La question « posée au test statistique » n'est pas tant celle de l'existence d'une liaison que celle de l'aptitude d'un ensemble de données aussi restreint à mettre en évidence cette liaison. De toutes façons, vu le petit nombre de données il convient d'être prudent dans l'interprétation (ce qui concrètement veut dire : retours fréquents aux données brutes et mise en relation avec des informations extérieures aux données).

### Les vins et les mots dans un même plan

L'analyse des correspondances est une méthode exploratoire qui fournit une représentation superposée des vins et des mots. La lecture de ce graphique peut être décomposée en trois étapes : 1) en regardant uniquement les positions des vins, deux vins sont proches sur le graphe s'ils sont caractérisés de la même façon par les dégustateurs, autrement dit par les mêmes mots. Deux vins sont éloignés s'ils sont caractérisés par des mots différents. 2) en regardant uniquement les mots, deux mots sont proches s'ils sont utilisés pour caractériser les mêmes vins et ils sont éloignés s'ils sont utilisés sur des vins différents. 3) en regardant simultanément les vins et les mots, on s'intéresse aux associations entre les vins et les mots. Globalement un vin apparaît du côté des mots auxquels il est souvent associé et éloigné des mots auquel il est rarement associé. Symétriquement, un mot apparaît du côté des vins auxquels il est souvent associé et à l'opposé des vins auquel il est rarement associé. Le plan de représentation proposée par l'analyse est optimal dans le sens où c'est le meilleur plan possible pour visualiser et résumer l'information contenue dans le jeu de données.

Le pourcentage d'information associé au plan montre dans quelle mesure celui-ci représente les données. Ici, le pourcentage est de 53.6% ; on peut considérer ce pourcentage comme élevé compte tenu du contexte : avec ce plan, on représente la moitié de la diversité des mots associés aux vins.



**Figure 1.** Représentation des mots et des vins fournie par l'analyse des correspondances du tableau 2.

Dans la représentation des vins (figure 1), on peut distinguer trois pôles :

Aubuisnières Silex (n°6), caractérisé par sucré, cité 11 fois pour ce vin; ce vin est le seul à contenir du sucre résiduel à une concentration nettement perceptible; cette caractéristique, inhabituelle (mais autorisée) dans un vin sec, ressort nettement en ce sens qu'elle est relativement peu citée pour les autres vins (7 fois mais jamais plus de 2 fois pour un même vin) et qu'elle constitue plus du tiers des mots associés à ce vin. Le graphique met plus en avant le faible nez de ce vin; mais ce terme n'étant cité que 3 fois pour ce vin, nous ne lui accordons qu'une seconde place (en outre, cette caractéristique est plutôt une absence de caractéristique, donc moins évocatrice; nous y reviendrons).

Aubuisnières Marigny (n°7) + Fontainerie Coteaux (n°10). Ces deux vins sont caractérisés principalement par bois, boisé, cités respectivement 7 et 5 fois pour chacun, ce mot n'étant utilisé que 3 fois par ailleurs. Cette description est évidemment à relier avec le fait que ces deux vins sont les seuls à avoir subi un élevage en fût. D'après le plan, goût étranger caractérise mieux ces vins, mais nous ne le citons qu'en second du fait de sa faible fréquence de citation (4), même si ce terme n'a été cité que pour ces deux vins. Remarquons au passage que l'effet de l'élevage en fût, dans l'ensemble plutôt recherché par la profession, n'engendre pas que des caractéristiques positives.

Les cinq Touraine (Sauvignon ; n° 1 à 5). Pour ces vins, les caractérisations sont moins évidentes. Citons bonne expression aromatique, frais, agrume, fin, discret qui sont en phase avec la classique image du Sauvignon, selon laquelle ce cépage conduit à des vins frais très aromatiques. Ajoutons à cela deux caractéristiques « en creux » : manque de fraîcheur (resp. peu de caractère...), cité uniquement pour les Chenins (8 fois ; resp. 4 fois).

Une fois ces trois pôles établis, on peut chercher à qualifier les axes. Le premier axe (horizontal), principale information du jeu de données, sépare les Sauvignons (à gauche) des Chenins (à droite) sur une base de fraîcheur et d'expression aromatique. Le second axe (vertical), information indépendante du premier axe, oppose les Chenins passés en fût (à goût boisé ; en haut) à celui présentant du sucre résiduel (au goût sucré ; en bas). Il différencie les Chenins entre eux, sans s'intéresser à une opposition Sauvignon – Chenin.

Ces grandes lignes étant établies, le mot nez faible utilisé pour les seuls vins 6 et 8 apparaît bien à sa place, c'est-à-dire loin des vins aromatiques, que ces arômes soient ceux du Sauvignon ou induits par le passage en fût.

Finalement, ce plan propose une image des vins blancs du Val de Loire selon laquelle les Sauvignons sont homogènes et les Chenins divers. Ce qui, in fine, pourra être traduit de différentes façons, non contradictoires :

il n'est qu'une seule façon de faire du Sauvignon et de nombreuses façons pour faire du Chenin ;

les viticulteurs « travaillent » plus leur Chenin, cépage blanc noble de la Touraine, en essayant divers itinéraires techniques.

Disant cela, nous sortons de notre rôle de statisticiens, mais nous avons voulu évoquer quelques façons dont l'utilisateur final peut s'approprier les résultats.

## Une image fiable

Il est ainsi possible d'obtenir, pour un ensemble de vins, une image fiable des associations entre ces vins et les mots qu'ils évoquent chez des goûteurs experts. La fiabilité est suggérée ici par les relations claires entre les « descriptions » sensorielles et les informations « externes » disponibles concernant le cépage et l'élevage (en fût ou non). Ce résultat est obtenu avec un protocole très allégé (une seule séance) comparé au protocole usuel, grâce à l'Analyse des Correspondances qui fournit une représentation plane de l'espace des vins, représentation assez facile à interpréter.

## Référence

1. Recherche initiée par C. Asselin, Interloire (Interprofession des Vins du Val de Loire) Angers et réalisée à Agrocampus, Rennes (école d'agronomie et d'agroalimentaire)

## Bibliographie

Escofier B. & Pagès J. (2011). Analyses factorielles simples et multiples : objectifs, méthodes et interprétation. Dunod, 4e édition, Paris.

Husson, F., Lê, S. & Pagès, J. (2009). Analyse de données avec R. Presses Universitaires de Rennes.

Lebart L., Piron M. & Morineau A. (2006). Statistique exploratoire multidimensionnelle: visualisation et inférence en fouilles de données. Dunod, 4e édition, Paris.