Corrélats textuels autour du concept de minéralité dans les vins

Pascale Deneulin[1,2,4], Laurent Gautier[2], Yves Le Fur[3], François Bavaud[4]

[1] Haute Ecole de Viticulture et Œnologie – Changins – pascale.deneulin@changins.ch

[2] Université de Bourgogne – laurent.gautier@u-bourgogne.fr

[3] AgroSup Dijon – yves.le-fur@agrosupdijon.fr

[4] Université de Lausanne – fbavaud@unil.ch

Abstract

Minerality has emerged as a widespread term in the language of professionals and wine consumers, yet lacking a precise and broadly shared definition. This contribution studies three parallel corpora, consisting of responses from 1697 consumers, underlining what minerality evokes (or not), how consumers define it and what terms can be considered as synonymous. Two methods are compared, namely Correspondence Analysis on one hand, highlighting the textual salience within each corpus, and clustering of textual networks generated by the *renormalized Markov Associativities* on the other hand, based on associations between terms. The two analyzes, complementary, distinguish and identify the various consumers' views regarding the concept of minerality in wines.

Résumé

Le terme *minéralité* a fait son apparition dans le discours des professionnels et des consommateurs de vins, sans qu'en existe une définition précise et consensuelle. Cette étude de trois sous-corpus parallèles (trois questions), constitués des réponses de 1 697 consommateurs, met en évidence ce que la minéralité évoque (ou non), comment les consommateurs la définissent et quels termes peuvent être considérés comme synonymes. On y compare les résultats obtenus par deux méthodes, l'Analyse des Correspondances d'une part, soulignant les saillances textuelles dans chacun des sous-corpus, et le partitionnement des réseaux textuels générés par les *Associativités Markoviennes renormalisées* d'autre part, basées sur les associations entre termes. Les deux analyses, complémentaires, distinguent et identifient les différentes représentations que les consommateurs peuvent avoir du concept de minéralité dans les vins.

Mots-clés : analyse des correspondances, analyse sensorielle, minéralité, modularité, Ncut, similarités entre termes, réseaux textuels, vin

1. Introduction : minéralité dans les vins et verbalisation

La visée de cet article est double, par son application aux domaines de la viticulture, de l'œnologie et de l'analyse sensorielle ainsi que par le développement d'outils méthodologiques. L'objectif du projet est de mieux cerner la notion de *minéralité* chez les consommateurs de vins (Gautier et al., 2014). Afin de dégager la structure lexicale contenue dans trois sous-corpus de parole, nous appliquons en premier lieu une Analyse des Correspondances classique sur chacun des tableaux lexicaux (section 2.3). Dans un second temps, nous étudions plus précisément les associativités entre mots au moyen des *associativités markoviennes* (section 3). Ces associativités permettent de générer un réseau textuel visualisant les similarités entre termes mais où les termes rares dominent. Afin de dépasser cette difficulté, des associativités renormalisées par le poids des termes sont définies et les termes sont regroupés en *communautés* maximisant la *modularité* de la partition. Cette alternative au *clustering* classique sur les distances du chi2 peut ainsi être éprouvée sur un

exemple concret de taille importante et sur trois corpus parallèles. Nous nous limiterons ici à l'analyse du vocabulaire utilisé et nous n'aborderons pas la segmentation des consommateurs qui fera l'objet d'une prochaine contribution.

2. Le questionnaire et les premières analyses

2.1. Constitution du corpus et prétraitements

Le corpus étudié est issu d'un questionnaire diffusé en ligne au moyen de plusieurs mailings en France et en Suisse francophone. Ainsi, 1 697 consommateurs y ont répondu entre 2011 et 2012. Soulignons qu'aucune autre étude textuelle porte à la fois sur un si grand nombre de réponses et sur la notion de minéralité dans les vins. La structure du questionnaire comporte deux parties : une première qui regroupe des éléments sociodémographiques incluant des questions sur les comportements d'achat et de consommation, une seconde composée de trois questions ouvertes. Afin d'appréhender sous différents angles les représentations que possèdent les répondants à l'égard *de la minéralité dans les vins*, les trois questions ont été formulées de la façon suivante :

(i) Si je vous parle de minéralité à propos de vin, à quoi cela vous fait-il penser ?

Imaginez que vous ayez à expliquer à un ami ce qu'est la minéralité d'un vin. Pour lui expliquer, (ii) vous donnez une définition, (iii) vous citez des synonymes.

Il a été choisi de rester très large dans la formulation des questions afin de ne pas limiter les réponses à un certain type de vin. La première question devait permettre aux répondants de s'exprimer le plus spontanément possible. La deuxième question étant plus encadrée, on peut supposer que les répondants s'efforcent alors de formuler une réponse la plus juste et plus précise en référence à leurs pré-acquis. Enfin, la dernière question a pour but de générer des synonymes et des exemples. Le questionnaire a été formaté de manière à ce que les répondants ne puissent pas laisser de champs vides. Au final, et après prétraitement textuel (cf. infra), on dispose donc de trois sous-corpus de parole, appelés par la suite évocation, définition et synonymes, regroupant chacun 1 697 réponses.

Chaque sous-corpus a été lemmatisé grâce au logiciel *TreeTagger* et un tableau lexical a été créé avec le logiciel *Textable* (Xanthos, 2014) sous *Orange Canvas*. Seuls les conjonctions (maintenues afin d'étudier les modes d'exemplarité associés à "comme", etc.), noms, noms propres, adjectifs et verbes ont été conservés. Les mots-outils restants ont été retirés manuellement; les réponses données sous forme de points de suspension ou de points d'interrogation ont cependant été conservées, et codées préalablement à la lemmatisation.

2.2. Taille et variété comparées des réponses lexicales

La table 1 résume et compare quelques caractéristiques globales des trois sous-corpus. La taille NN mesure le nombre de termes (*tokens*), et la variété V le nombre de termes distincts (types). $H = -\sum_j f_j \ln f_j$ est l'entropie en nats, où f_j est la fréquence relative du *j-ème* terme, cumulée sur tous les répondants. *Hapax* donne le nombre de termes apparaissant une seule fois, $GT = hapax/V$ est l'estimateur de Good-Toulmin de la probabilité que, si le questionnaire avait été étendu à un répondant supplémentaire, le premier terme émis ait été inédit (Bunge et Fitzpatrick, 1993). *Pente* est l'estimation des moindres carrés de β dans la loi de Zipf $\ln f_{[i]} = c - \beta\, ln_{[i]}$, où [i] est le rang moyen décroissant du terme i; *pente pondérée* est l'estimation de ce même paramètre par moindres carrés pondérés par les fréquences relatives $f_{[i]}$.

	N	V	H	hapax	GT	pente	pente pondérée
évocation	14'293	1'667	5.53	838	0.50	1.17	0.99
définition	15'488	1'691	5.52	812	0.48	1.20	0.99
synonymes	5'156	1’008	5.55	582	0.58	1.04	0.92

Table 1. Taille et richesse comparées des réponses lexicales lemmatisées

Malgré l'abondance des études consacrées au sujet, la comparaison de la richesse lexicale de corpus de tailles différentes demeure un exercice périlleux. Toutefois, le sous-corpus évocation apparaît plus riche que le sous-corpus définition, pourtant de taille comparable, au sens où la raréfaction des termes moins fréquents est moins rapide dans évocation que dans définition, selon les indicateurs *hapax*, *pente* et *pente pondérée.* Aussi, l'entropie H (ainsi que *GT*), mesurant l'incertitude sur les termes au sens de la Théorie de l'Information, est plus grande pour évocation que pour définition, malgré une variété V inférieure - un résultat attendu au vu de la différence des tailles N. Le corpus synonymes est le moins varié au sens de V, mais paradoxalement le plus riche au sens des autres indicateurs.

2.3. Analyses des Correspondances

Afin de dégager la structure lexicale des trois sous-corpus en *relation avec le concept de minéralité*, des Analyses des Correspondances (AC) sont effectuées avec *FactoMineR* (Lê et al., 2008) sur chacun des 3 tableaux lexicaux “répondants fois termes”, en ne conservant que les termes d'effectif de 5 citations minimum. De façon générale, les premiers facteurs obtenus sont créés par des termes *n'expliquant en rien* la minéralité dans les vins. Ces termes seront dans un deuxième temps éliminés afin de dégager les oppositions exclusivement associées au concept étudié.

2.3.1. Sous-corpus évocation

Les douze premières dimensions de l'analyse des correspondances mettent en évidence des points particulièrement périphériques, très éloignés du profil moyen. Tous ces termes traduisent *l'indétermination* de ce qu'est la minéralité. Ainsi, nous retrouvons les points d'interrogation sur l'axe 1 (2.11%), les termes “rien, précis, désolé“ sur l'axe 2 (1.86%), “idée” pour les réponses comme “*je n'en ai aucune idée*” sur l'axe 3 (1.33%), “grand-chose” pour “*pas grand-chose*“ sur l'axe 4 (1.14%), “savoir, tout” pour des réponses comme “*je ne sais pas du tout ce que c'est*” sur l'axe 5 (1.13%), “désolé, précis” pour des réponses “*désolé*, *rien de précis*” sur l'axe 6 (0.93%), aucun terme particulier sur l'axe 7 (0.90%), “tout” sur l'axe 8 (0.82%), aucun terme particulier sur l'axe 9 (0.74%), “avis et vignoble“ en opposition sur l'axe 10 (0.73%), “vignoble” sur l'axe 11 (0.71%), “connaître” pour “*Je ne connais pas ce terme*” sur l'axe 12 (0.68%).

Souhaitant avant tout analyser ici les sens sous-jacents du terme minéralité pour les consommateurs, nous avons exclu ces onze termes n'apportant pas d'information quant à l’évocation de la minéralité. Une seconde analyse des correspondances a été réalisée sur le nouveau tableau lexical réduit comportant V = 372 termes et 1 562 répondants possédant des réponses réduites non vides. Les 135 répondants écartés se subdivisent de la façon suivante: 13 répondants qui utilisent un vocabulaire trop spécifique et dont la fréquence de citation est inférieure à 5 (répondants enlevés dès le premier tri de termes basé sur les fréquences de citation) et 122 répondants qui n’ont pas été en mesure de formuler ce à quoi leur fait penser la minéralité dans les vins.

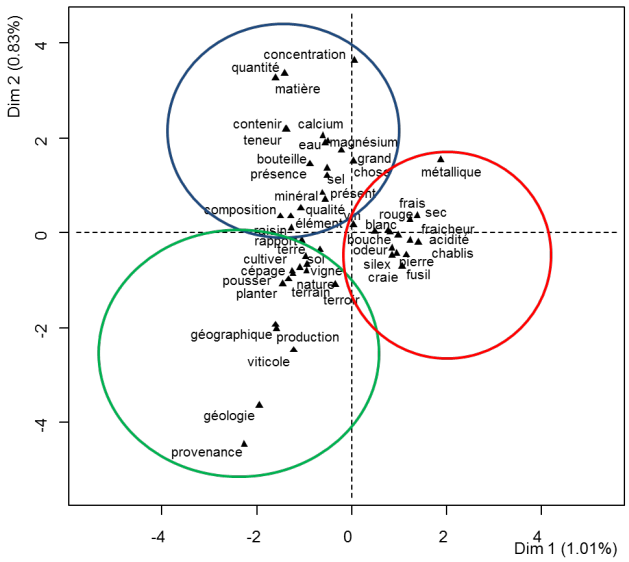

Figure 1. Sous-corpus évocation : Analyse des correspondances, après élimination des termes saturant sur les premiers facteurs et non reliés au concept de minéralité

Afin de rendre le graphique plus lisible, seuls les 50 termes possédant les *cosinus carrés* les plus élevés sur les deux premiers axes ont été conservés sur la figure. Il s'agit donc des 50 termes les mieux représentés sur ce premier plan (figure 1). Le premier axe met en évidence des termes relevant à la fois du champ lexical du sensoriel comme “bouche, odeur, acidité, fraîcheur” mais aussi “pierre, fusil et silex” (pris ici comme attributs sensoriels) ou relevant de l’exemplification comme “Chablis” ou plus généralement vin “blanc sec” : “*Plutôt aux vins blancs; secs; avec arômes de pierre à fusil”. “Des arômes, dans le vin, pierre à fusil, silex, allumettes, pas trop d'acidité, vin blanc exemple : Chablis”.*

Le deuxième axe met en évidence deux autres champs lexicaux qui s'opposent entre eux : le champ lexical associé aux minéraux et aux eaux minérales avec les termes “eau, quantité, teneur, minéral, sel, bouteille, calcium, magnésium” et à l'opposé, un lexique géo-pédoclimatique et viticole faisant référence au sol, au terroir et à la vigne avec les termes “provenance, géologie, sol, terrain, terroir, vigne, cépage, pousser, planter”. *“Minéraux : calcium, potassium, sodium, magnésium pureté, légèreté noter le côté “salé” du vin et non l'acidité”. “Difficile à qualifier : le rapport avec le terroir où est cultivée la vigne; rapport avec le sol et les roches.”*

L'étude de ce premier sous-corpus montre que les principales saillances portent avant tout sur l’absence de pré-acquis à l’égard de la minéralité : une part importante des consommateurs n'est pas capable de dire à quoi lui fait penser la minéralité, tandis que pour une autre part, la notion se scinde en trois domaines de références :

1) le lien au champ lexical sensoriel sous-tendu par pierre à fusil et silex, illustré par des exemples de vins blancs,
2) le lien aux eaux minérales et à leurs sels minéraux,
3) le lien au sol, au sous-sol et au terroir.

2.3.2. Sous-corpus définition

Le sous-corpus définition est de taille comparable au précédent. Les huit premières dimensions de l'analyse des correspondances traduisent également le *non savoir* de ce qu'est la minéralité.

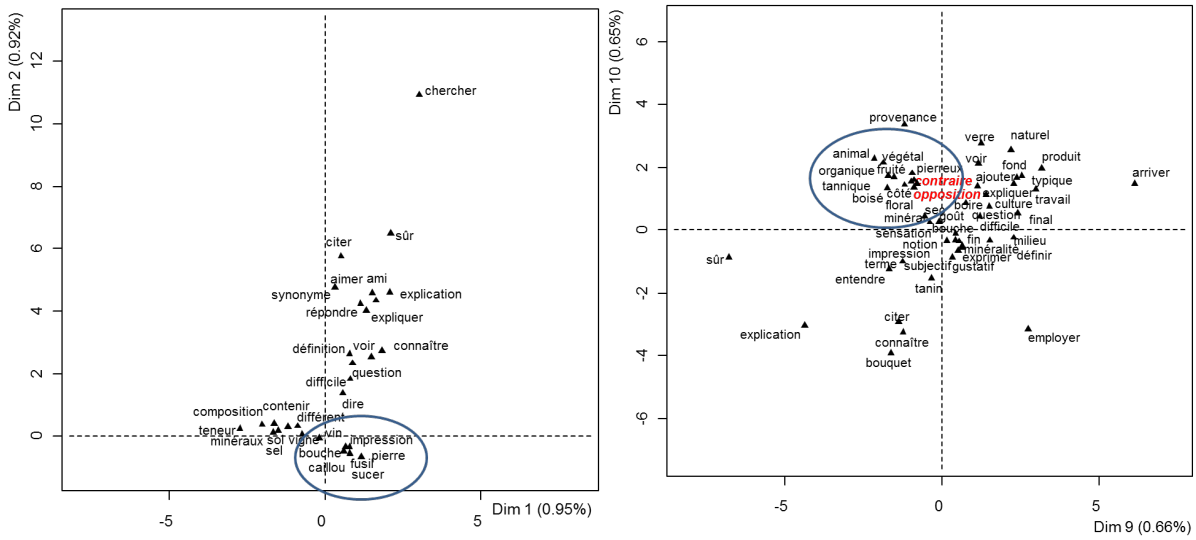

Figure 2. AC sur le sous-corpus réduit définition*, facteurs 1 et 2 (gauche); facteurs 9 et 10 (droite)*

Après suppression des termes les mieux associés aux 7 premiers facteurs (points de suspension et d'interrogation, idée, savoir, être, incapable, réponse, précédent), une nouvelle AC a été effectuée sur le tableau lexical réduit contenant V = 378 termes et 1 478 répondants. Il s'avère que 35 répondants utilisent un vocabulaire peu cité (inférieur à 5 citations) et 184 répondants ne sont pas en mesure de donner une définition de la minéralité. L'étude des 30 termes les plus corrélés aux deux premiers axes (figure 2 gauche) est difficile à interpréter. Les termes corrélés à l'axe 2 suggèrent encore des traces d'indétermination par l'emploi de verbes comme "chercher, expliquer, citer, répondre". Nous pouvons toutefois remarquer une association particulièrement intéressante entre les termes "bouche, impression, sucer, caillou, pierre" : "*La minéralité peut être assimilée à l'impression de sucer un caillou.*".

L'étude des axes suivants permet de retrouver des proximités lexicales instructives sur le plan formé par les axes 9 et 10 (figure 2 droite). Nous y retrouvons les familles aromatiques que sont "fruité, végétal, animal, boisé et floral" à côté des termes "opposition et contraire". A défaut de pouvoir donner une définition précise, les consommateurs pratiquent l'usage des contre-exemples : *"Lien avec le minéral bien sûr. En opposition avec ce qui est végétal."*.

Il apparait donc nettement plus exigeant pour les consommateurs de donner une définition de la minéralité (184 répondants en sont incapables) que d'exprimer ce que la notion leur évoque à propos de vin (corpus évocation). La définition par le contraire est un recours : à défaut de définir *ce qu'est la minéralité*, les consommateurs se réfugient derrière des formes de définition fondées sur *ce que la minéralité n'est pas*.

2.3.3. Sous-corpus synonymes

Le sous-corpus synonymes est de taille nettement inférieure aux deux corpus précédents mais révèle des structures lexicales similaires. Les six premières dimensions de l'analyse des correspondances sont entièrement corrélées aux différentes formes d'indétermination.

Après suppression des 10 termes associés à ces 6 premiers facteurs, une nouvelle AC est effectuée sur le tableau lexical contenant V = 162 termes et 1282 répondants. Là encore, 90 répondants utilisent un vocabulaire trop spécifique (nombre de citations inférieur à 5) et 325 répondants ne parviennent pas à donner des synonymes à la minéralité.

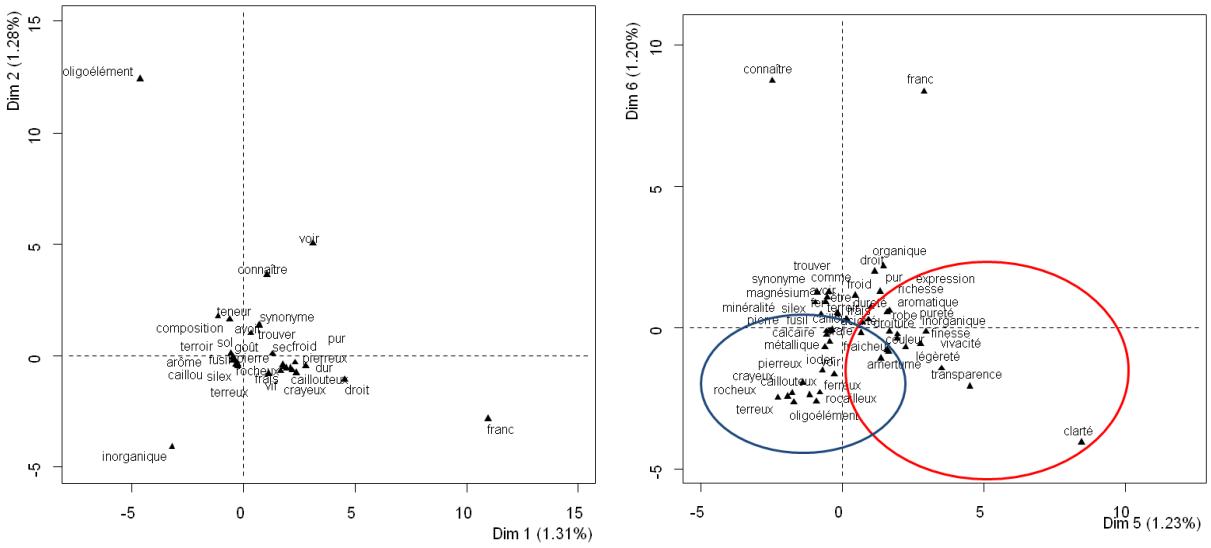

Figure 3. AC sur le sous-corpus réduit synonymes*, facteurs 1 et 2 (gauche); facteurs 5 et 6 (droite)*

Le premier plan de l'analyse des correspondances (figure 3 gauche) met en évidence trois termes fortement corrélés aux axes 1 ou 2. “franc” et “inorganique” s'opposent sur l'axe 1 tandis que “oligoélément” ressort sur l'axe 2. L'étude du plan 5-6 (figure 3 droite) montre des regroupements particuliers : un premier regroupement autour du champ lexical de la sensorialité (en rouge) où figurent “clarté, fraîcheur, finesse, amertume, vivacité” et un autre regroupement qui s’oppose au premier dans le domaine pédo-géologique (en bleu) avec notamment “pierreux, caillouteux, rocheux, terreux, crayeux, oligoélément”.

3. Association entre termes et réseaux lexicaux

3.1. Associativités markoviennes simples et renormalisées

Un sujet répondant i = “fût” a davantage de chances de répondre également j = “chêne” que n'en a le répondant typique muni du profil lexical moyen : “fût” et “chêne” sont alors associés. L'*associativité markovienne* s_{ij} (Bavaud et Xanthos, 2005) mesure le rapport entre la probabilité jointe de cooccurrence de la paire ij et sa valeur attendue sous indépendance. Toujours symétrique, elle caractérise l'attractivité de la paire lorsque s_{ij}>1, et sa répulsivité lorsque $0 \leq s_{ij} < 1$. Parmi les paires maximalement similaires dans le corpus évocation, on trouve ainsi “point-vue”, “croire-dépôt”, “eeeuh-désolé”, etc.

Les premières expérimentations montrent que les similarités $S = (s_{ij})$ tendent à être d'autant plus grandes que les termes en jeu sont rares, i.e. tels que leurs fréquences relatives f_i, f_j sont petites. Pour construire un graphique textuel portant sur des paires de termes fortement associés et suffisamment fréquents, nous considérerons dans cette contribution les *similarités renormalisées*

$$T = (t_{ij}) \qquad \text{où} \qquad t_{ij} := f_i^{\alpha} f_j^{\alpha} s_{ij}$$

où le paramètre de renormalisation $\alpha \geq 0$ permet de surpondérer la contribution des termes suffisamment fréquents. Concrètement, on sélectionne les P paires *distinctes* ij pour lesquelles t_{ij} est maximal, ce qui aboutit à la sélection de $n(P)$ termes formant les nœuds d'un réseau lexical. De plus, les cooccurrences entre termes exprimées par les associativités

markoviennes peuvent, à l'instar d'une chaîne de Markov, être itérées (co-co-occurrences etc.), et définir ainsi des *similarités d'ordre r*.

3.2. Comparaison des paramètres de renormalisation sur le sous-corpus évocation

La figure 4, obtenue avec le logiciel *Gephi*, visualise les similarités entre les P=100 premières paires du corpus évocation maximisant les associativités renormalisées $t_{ij}^{(r)}(\alpha)$, où α est le paramètre de renormalisation et r dénote l'ordre de la chaîne. La taille des arrêtes est donnée par les *associativités renormalisées*, la taille des nœuds par la *centralité d'intermédiarité* du réseau associé (Brandes, 2001), et leur couleur code leur groupe tel que résultant d'une détection de communauté maximisant la *modularité* du réseau, selon l'algorithme itératif d'agrégation de Blondel et al. (2008), aboutissant à $c(P)$ communautés (table 2).

Le recours au critère de modularité maximale (Girvan et Newman, 2002) peut être comparé au critère de minimisation du *Ncut* (Shi et Malik, 2000) : tous deux visent à partitionner un réseau pondéré de termes en groupes aussi « intra-connectés » et « inter-isolés » que possible. Des considérations mathématiques, dépassant malheureusement le cadre de cette contribution, témoignent d'une part de la proximité formelle du *Ncut* avec le *spectral clustering* et l'*analyse des correspondances classique* (section 2.3); d'autre part, on peut montrer que l'application du critère de modularité sur les associativités renormalisées (figures 4 et 5) constitue en première approximation une variante de l'analyse des correspondances, sous-pondérant les termes fréquents pour α<0.5, et les surpondérant pour α>0.5.

ordre r	paramètre α	nombre de termes sélectionnés $n(P)$	nombre de communautés $c(P)$
1	0.5	136	46
1	0.8	41	6
1	1	33	5
2	0.5	53	12
2	0.8	27	4
2	1	22	2

Table 2. Caractéristiques des réseaux associatifs lexicaux du corpus évocation, selon α et r, pour P=100

Tant $n(P)$ que $c(P)$ décroissent avec α (pour r fixé) et avec r (pour α fixé). D'un point de vue interprétatif, on souhaiterait conserver suffisamment de termes, pour autant qu'ils soient fortement associés entre eux au sein de groupes peu nombreux, afin d'obtenir une représentation conservant un maximum d'informations tout en restant lisible. Les réseaux (r=1, α=0.8) et (r=2, α=0.5) de la figure 4 semblent constituer les meilleurs compromis à cet égard.

Le réseau textuel r=1, α=0.8 (figure 4, en haut au centre) souligne l'importance des termes "minéral, pierre, goût, sol et terroir", tant par le nombre de connexions établies, entre 12 pour terroir et 19 pour minéral, que par la centralité d'intermédiarité du réseau associé, c'est-à-dire le nombre de plus courts chemins passant par ces termes-sommets. Les associations les plus fortes sont entre "faire" et "penser", suivie de "fusil" et "pierre", puis par "goût" et "pierre".

Les 6 communautés détectées montrent que la notion de minéralité à propos de vin évoque :

1) en rouge : une eau minérale, sa teneur et sa composition en sel. Les consommateurs raisonnent alors par analogie. Ils font référence à ce qui leur vient le plus spontanément à l'esprit à propos de minéralité : celle de l'eau minérale. A titre d'exemple, on retrouve ici : *"goût calcaire comme une eau minérale style Contrex"*,

"cela évoque la présence, dans le vin (au même titre que dans l'eau minérale), d'éléments minéraux provenant du sol et donc attachés à la qualité et aux spécificités du terroir ... "

2) en violet : un "goût, un arôme ou une sensation" qui fait écho au "terroir, à la terre ou au cépage". Cette combinatoire entre univers sensoriel, géo-pédologique et cultural est ici omniprésente. Elle se traduit par des formulations plus ou moins élaborées, parfois lapidaires : *"pour moi, la minéralité d'un vin est associée au terroir qui donne l'arôme, le goût, la robe ..."* puis glisse directement sur la sémantique sensorielle : *"au goût de terroir du vin, à l'odeur", "le goût en bouche, le goût du terroir"*.
3) en vert : au "sol, au terrain où pousse la vigne". On retrouve là-encore la notion de terroir, mais limité plus spécifiquement au sol et à la nature du terrain *"Terrain sur lequel pousse la vigne - Types de roches concernés."*.
4) en bleu clair : une "odeur de pierre à fusil, de silex, de caillou, de calcaire, un blanc sec, un Chablis". On retrouve ici une évocation répandue, celle de l'odeur de la pierre à fusil, de silex entrechoqués, doublée d'une volonté d'exemplifier à travers le cas des vins blancs, ceux de Chablis notamment. *"ça évoque un goût de caillou, de pierre à fusil. Typique des vins blancs de chablis"*.
5), 6) "(pas) grand-chose" ; "rien, désolé"

Le réseau textuel d'ordre r=2 (les voisins des voisins), α=0.5 (figure 4, en bas à gauche) montre une centralité élevée pour les termes "minéral, goût, pierre, blanc, terroir, vigne", allant de 37.8 pour vigne à 169.5 pour minéral. En particulier, l'importance du terme "blanc" (centralité de 57.7) est soulignée dans cette représentation d'ordre r=2, plusieurs liaisons telles qu'entre "sec" et "Chablis" ne pouvant transiter que par ce sommet, par ailleurs fortement associé à "vin" (préalablement éliminé de l'analyse). Les paires les plus fortement associées sont "désolé – rien", puis "savoir – tout" ("*je ne sais pas du tout*"), puis "rien – précis". Ce réseau textuel fait donc aussi ressortir les associations relatives à ces formes d'indétermination. Finalement, 12 communautés sont ici formées des communautés déjà évoquées à propos du réseau textuel r=1, α=0.8, auxquels s'ajoutent les binômes suivants :

- en bleu foncé : "métallique – froid" comme dans *"(...) C'est une notion qui est pour moi difficilement accessible... Mais j'imagine quelque chose de sec, froid, quasiment métallique"* : le terme minéralité n'aurait donc pas toujours une connotation positive.
- "provenance - géographique".
- "ajouter – soufre" : plusieurs personnes font référence "*au soufre ajouté dans le vin*".
- "clarté – légèreté" comme dans "*La légèreté du vin en bouche et la clarté de sa robe"*.
- "liquide – robe" ainsi que "jeune – bourgogne" et "contenu – taux".

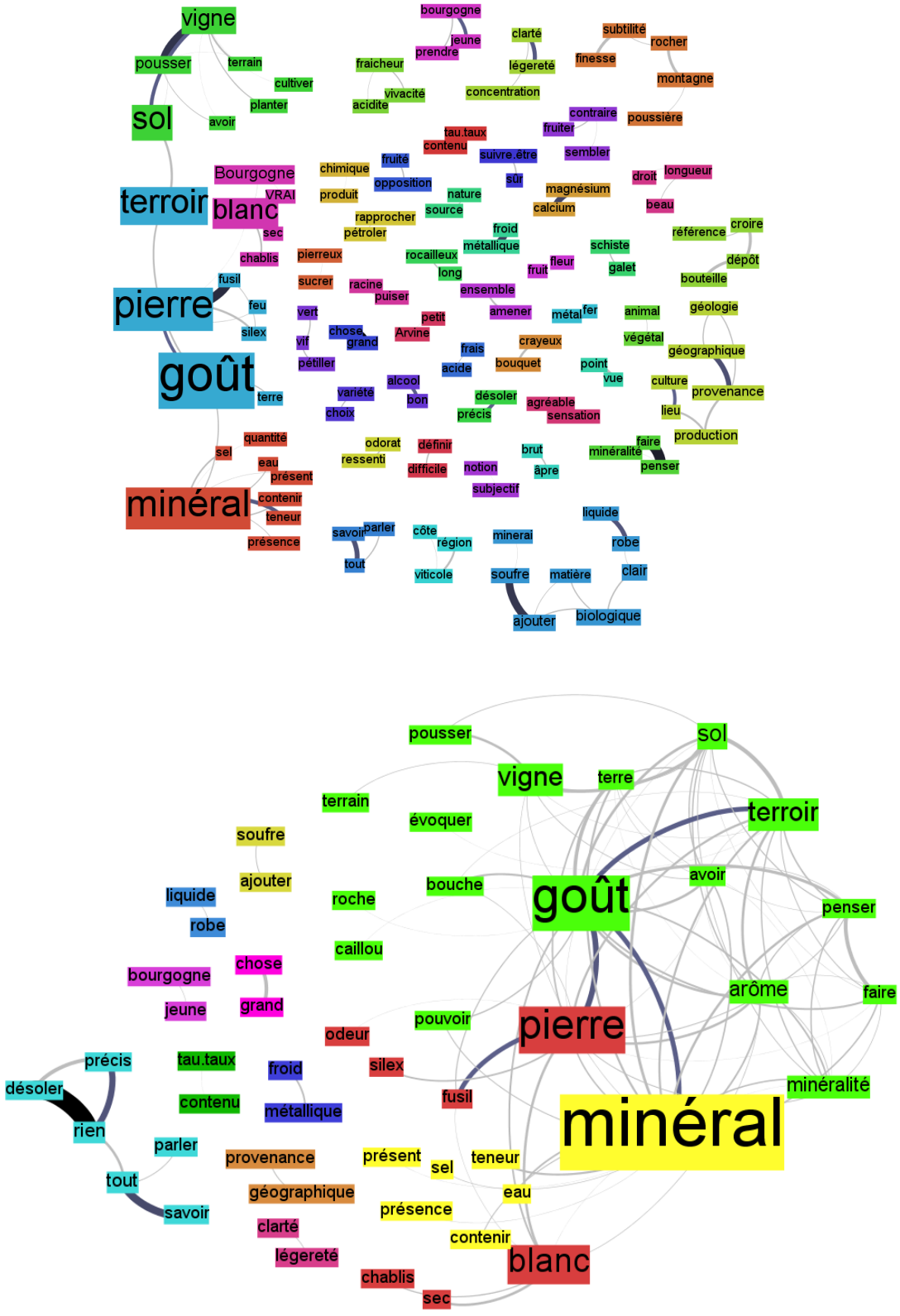

Figure 4. Réseau des associativités markoviennes renormalisées du corpus évocation*, de paramètres α=0.5, d'ordre r=1 (cooccurrences lexicales, en haut) et r=2 (co-co-occurrences, en bas)*

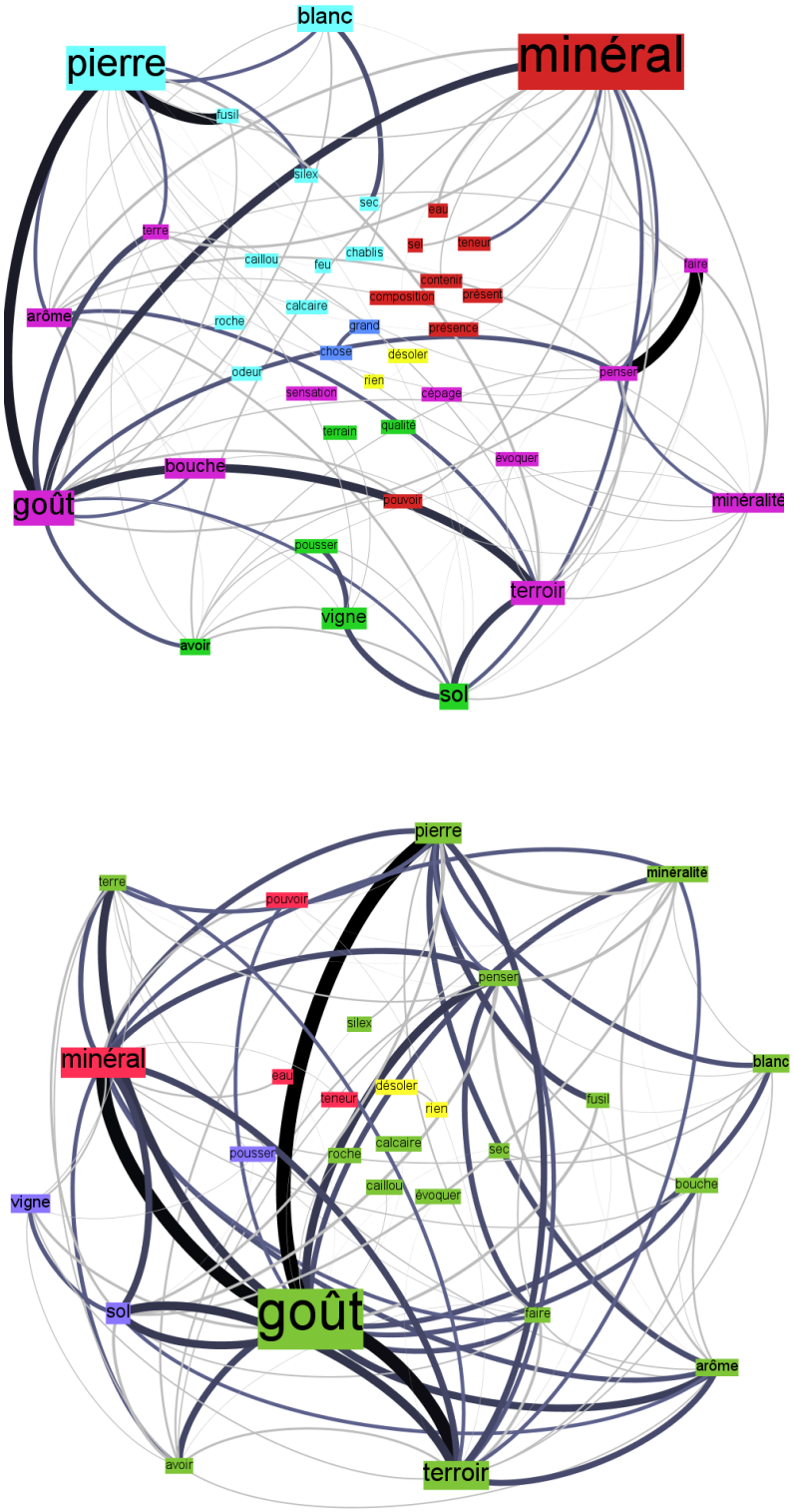

Figure 5. Réseau des associativités markoviennes renormalisées du corpus évocation*, de paramètres α=0.8, d'ordre r=1 (cooccurrences lexicales, en haut) et r=2 (co-co-occurrences, en bas)*

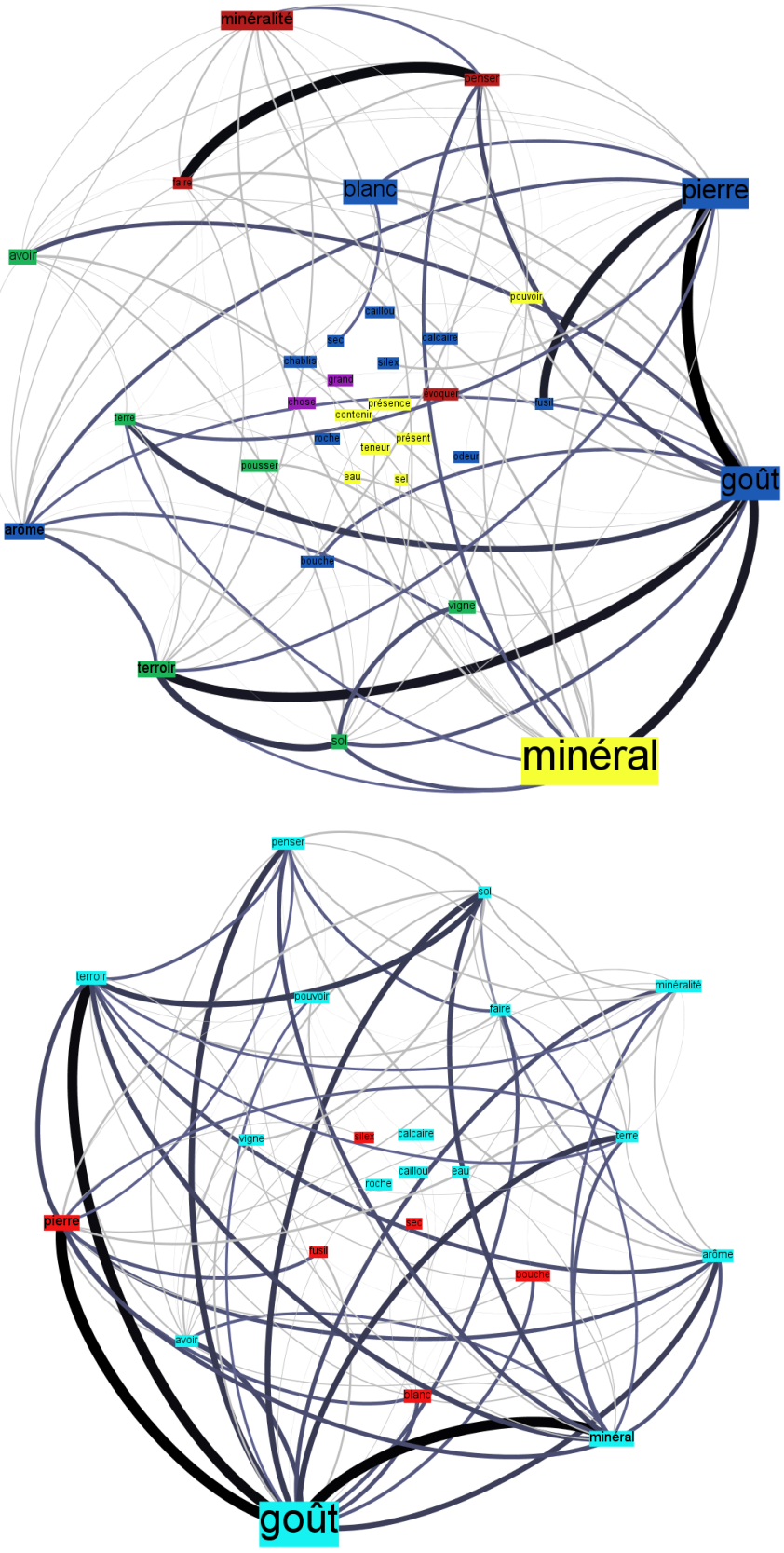

Figure 6. Réseau des associativités markoviennes renormalisées du corpus évocation, de paramètres α=1), d'ordre r=1 (cooccurrences lexicales, en haut) et r=2 (co-co-occurrences, en bas)

3.3. Sous-corpus définition et synonyme

Dans le réseau textuel du sous-corpus définition pour r=1 et α=0.8 (figure 5, gauche), les termes de plus forte centralité sont "goût, minéraux et minéralité". Les associations les plus fortes, en terme d'associativités markoviennes renormalisées, se tissent à partir du terme "goût", avec dans l'ordre : "un goût de pierre", "avoir un goût", "un goût de terroir", "un goût de minéralité", "un goût de terre" : la définition de la minéralité relève à l'évidence du *goût*. Les communautés de termes les plus saillantes sont ici :

- en rouge : une communauté liée à la sphère olfactive avec "arôme et odeur" associés à "pierre, fusil et silex".
- en jaune : autour du terme *goût,* une nouvelle communauté se dessine à partir des termes "sensation, sentir, ressentir, bouche, sucer et caillou". On y trouve par exemple : "*En bouche, c'est l'impression de sucer un caillou...*"
- en rose : l'expression du "terroir, du sol et de l'acidité". Le terroir est ici plus spécifiquement associé au sol mais également à l'acidité des vins blancs secs : *"vin blanc d'un terroir particulier, ayant un arôme équilibré plutôt acide, frais, léger"*.
- en vert clair : les références aux sels minéraux, à leur quantité et leur composition, sans allusion particulière à l'eau minérale. On retrouve ici des formulations comme : *"La minéralité est la composition du vin en minéraux"*, *"Ce sont les arômes qui sont proches des odeurs que l'on retrouve dans les minéraux. (...)"*.
- "fait penser" et "être incapable" (de définir).

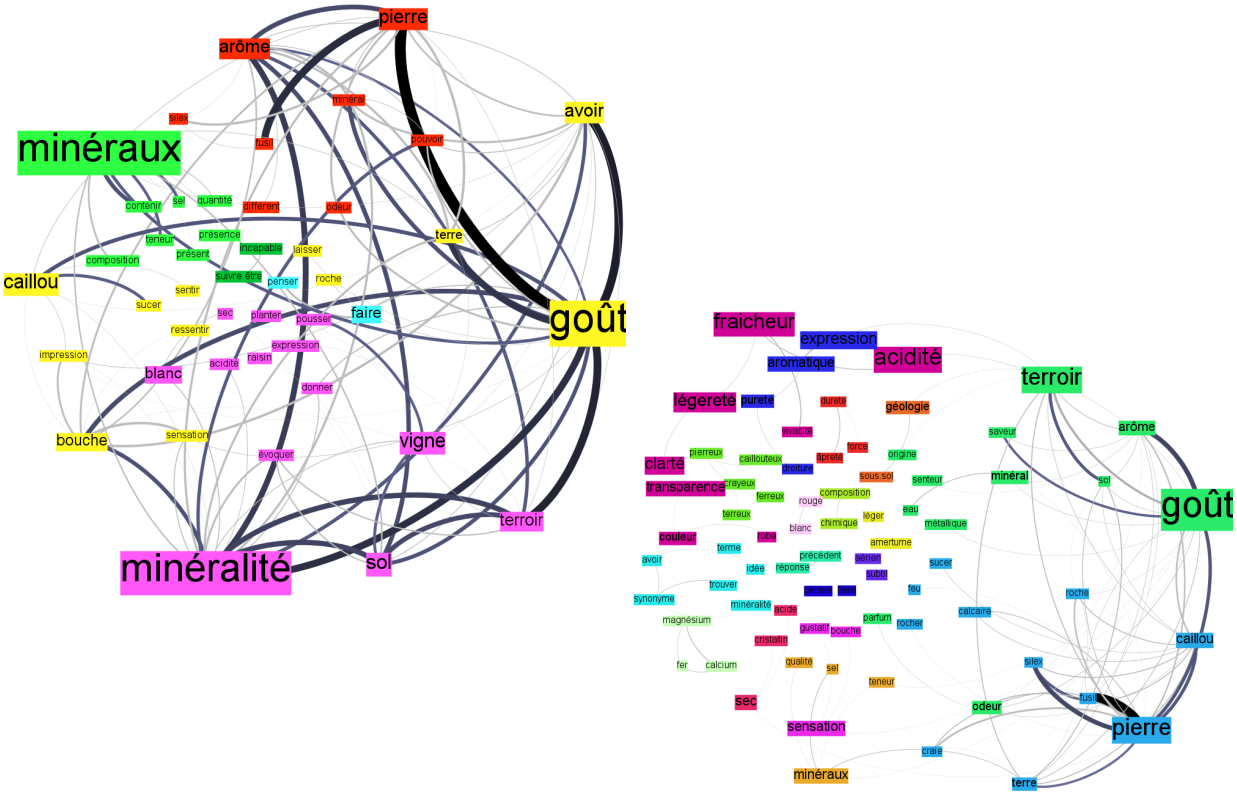

Figure 7. Réseau des associativités markoviennes renormalisées des sous-corpus définition *(gauche) et* synonymes *(droite), pour r=1 et α=0.8*

L'étude du sous-corpus synonyme pour r=1 et α=0.8 est riche en vocabulaire avec 76 termes représentés (pour 100 associations étudiées) et 17 communautés mises en évidence. Cette richesse s'explique par le fait que les consommateurs ont le plus souvent cité une liste de termes sans construire de véritables phrases (section 2.2). Elle entraîne une moindre connectivité du réseau textuel (figure 5, droite). Les termes les plus centraux sont, dans l'ordre, "goût, pierre, terroir, acidité et fraîcheur". Les 18 communautés caractérisent ici plusieurs thématiques :

- Les deux communautés avec les réseaux les plus denses sont associées à la notion de terroir. Ces deux communautés sont celles qui associent d'une part "goût de terroir, arôme, lié à l'origine, à l'eau minérale" en vert à droite et d'autre part "géologie et sous-sol" en orange. Ici, les notions de goût de terroir et d'eau minérale se trouvent fusionnées au sein d'une même communauté.
- en bleu : le lexique du minéral semble se départir du sensoriel : aucun des termes "pierre à fusil, caillou, roche, calcaire, silex, terre, rocher, craie" n'est fortement lié à "goût", "odeur" ou "arôme".
- en vert (à gauche) : une autre communauté lexicale constituée des adjectifs : "rocheux, caillouteux, crayeux, pierreux et ferreux" qui s'écarte du registre sensoriel.
- deux communautés relatives aux sels minéraux, l'une (en brun) où le terme "minéraux" est associé à "sel, teneur et qualité", l'autre (en vert clair) où les sels sont explicitement cités : "Fer, calcium et magnésium".
- Le binôme "composition - chimique" (en vert au centre).
- six communautés sont associées à des notions relevant du lexique sensoriel : "sensation gustative (en) bouche" pour une notion imprécise ; "dureté, âpreté, force" pour un champ lexical à connotation plutôt négative ; "subtil, aérien" pour un champ lexical de la finesse et la légèreté ; "acide, sec et cristallin" qui mélangent les perceptions gustatives et visuelles.
- en rose, la fraicheur en bouche "fraîcheur, acidité, vivacité, légèreté, clarté, transparence, couleur, robe", et en bleu foncé le champ lexical du zéro défaut avec "droiture, pureté de l'expression aromatique".
- "(vin) blanc" et "(vin) rouge" pour ceux qui différencient la minéralité dans les vins blancs et dans les vins rouges "*métallique pour un blanc, ou terreux pour un rouge*".
- "faire penser" mais aussi ceux qui ne savent pas : "synonyme, avoir, idée, trouver", ou ceux qui se réfèrent à la "question précédente".

4. Conclusion

De manière générale, l'ensemble des trois sous-corpus a d'abord mis en évidence les formes lexicales citées par les consommateurs qui *ne connaissent pas la minéralité*, ceux incapables d'apporter une réponse précise. Ce nombre est le plus grand pour la question synonymes, intermédiaire pour la question définition, et moindre pour la question évocation, où, laissant libre court à leur imagination, les consommateurs font ressortir l'association du concept de minéralité :

1) avec la notion de terroir, soit directement à travers le goût de terroir soit à travers les éléments constitutifs du terroir qui influent sur le goût final du vin,
2) avec le binôme sol-vigne,
3) avec les odeurs de pierre à fusil et de silex caractéristiques de l'appellation Chablis,
4) avec les eaux minérales et leur composition en minéraux.

Ces représentations apparaissent également dans les deux autres corpus. La question définition, bien qu'apparemment plus exigeante cognitivement pour les consommateurs, met en évidence deux représentations supplémentaires, à savoir l'exemplification avec “comme” “sucer un caillou” et l'utilisation de l'antagonisme. Enfin, l'étude de la question synonyme fait apparaître en plus un important champ lexical lié à la sensorialité, le plus souvent avec des descripteurs à connotation positive, bien que quelques descripteurs négatifs fassent aussi leur apparition.

Les deux méthodes statistiques utilisées sont complémentaires. Les principales différences observées reposent sur la mise en évidence de l'utilisation du contre-exemple par l'analyse des correspondances (sous-corpus définition) et une plus grande précision avec les *associativités markoviennes*, ou en tout cas une lecture plus facile des termes sensoriels présents dans le corpus synonymes. L'analyse des correspondances cherche à mettre en évidence les termes qui s'éloignent le plus du profil moyen, c'est-à-dire qui sont les plus saillants, qu'ils soient utilisés de manière isolée ou en association avec d'autres termes. Par contraste, les associativités markoviennes, renormalisées ou non, considèrent avant tout les associations entre termes plutôt que les saillances individuelles, et permettent de générer de nouveaux réseaux textuels en itérant à volonté les voisinages lexicaux. Cela étant, des arguments formels tendent à rapprocher la méthode basée sur la modularité des associativités renormalisées de la section 3 d'une variante de l'analyse des correspondances classique, surpondérant toutefois les termes fréquents pour $\alpha>0.5$, comme illustré dans cette contribution. Des analyses supplémentaires, tant pratiques que théoriques, seraient naturellement requises pour étayer cette thèse, méthodologiquement importante, et propre à contribuer à l'unification de l'approche classique de l'analyse des correspondances avec les apports plus récents des travaux sur les réseaux.

Remerciements

Les auteurs remercient l’Institut Œnologique de Champagne pour avoir initié cette étude sur la minéralité. Les données de ce papier font partie d’un projet sélectionné dans le cadre du programme de coopération territoriale européenne Interreg IV France-Suisse.

Références

Bavaud F. et Xanthos A. (2005). Markov associativities. *Journal of Quantitative Linguistics,* 12, pp. 123-137.

Blondel V.D., Guillaume J.-L., Lambiotte R. et Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics,* 10, P10008.

Brandes U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology,* 25, pp. 163-177.

J. Bunge J. et Fitzpatrick M. (1993). Estimating the number of species: a review. *Journal of the American Statistical Association,* 88, pp. 364-373.

Gautier L., Le Fur Y. et Robillard B. (2013). Les descripteurs du vin en Europe: regards contrastifs (= InnTrans). In Laurent Gautier \& Eva Lavric (Eds), La Minéralité du vin: mots d'experts et de consommateurs. Frankfurt/Main, etc.: Peter Lang.

Lê S., Josse J. et Husson F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software,* 25, pp. 1-18.

Girvan M. et Newman M. E. J. (2002). Community structure in social and biological networks. *Proceeding of the National Academy of Sciences of United States of America,* 99, pp.7821-7826.

Shi J. et Malik J. (2000). Normalized Cuts and Image Segmentation. *IEEE transactions on pattern analysis and machine intelligence,* 22, pp. 888-905.

Xanthos A. (à paraître). Textable: programmation visuelle pour l'analyse de données textuelles. In *Actes des 12èmes Journées internationales d'analyse statistique des données textuelles (JADT 2014)*.