





# **Combining statistics and semantic for an automated data** analysis of Free-Comment sensory description of products

Michel Visalli, Benjamin Mahieu

michel.visalli@inrae.fr

benjamin.mahieu@oniris-nantes.fr



15th Pangborn Sensory Science Symposium

Meeting new challenges in a changing world



20 - 24August 2023 Nantes France



https://www.pangbornsymposium.com/

# **Overview**

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

Consumer	Product	Description
C 1	D 1	good taste, good vegetable smell but
	C_1 P_1	too compact
		•••
		strong taste, nice color, vegetables are
C_N <sub>C</sub>	P_N <sub>P</sub>	varied

Consumer	Product	Description	Instructions (facultative)
C_1	P_1	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)
		•••	
C_N <sub>c</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied	Quality vs. Defect

Consumer	Product	Description	Instructions (facultative)
C_1	P_1	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)
		•••	
C_N <sub>c</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied	Quality vs. Defect

Consumer	Product	D_1	D_2	D_2		D_N <sub>D</sub>	Instructions (facultative)	
C_1	P_1	1	1	1		0	Sensory modality (visual aspect, texture, flavor, etc.)	
							Quality vs. Defect	
C_N <sub>c</sub>	P_N <sub>P</sub>	1	0	1	•••	0		

Consumer	Product	Description	Instructions (facultative)
C_1	P_1	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)
		•••	
C_N <sub>c</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied	Quality vs. Defect

Consumer	Product	D_1	D_2	D_2		D_N <sub>D</sub>	Instructions (facultative)	
C_1	P_1	1	1	1	•••	0	Sensory modality (visual aspect, texture, flavor, etc.)	
		•••	•••	•••	•••	•••	Quality va Dafaat	
C_N <sub>c</sub>	P_N <sub>P</sub>	1	0	1	•••	0	Quality vs. Defect	
Descriptor cited or not								

Consumer	Product		D	escriptio	n		Instructions (facultative)
C_1	P_1	good	. 0	od veget		ell but	Sensory modality (visual aspect,
			to	o compa	ct		texture, flavor, etc.)
C_N <sub>C</sub>	P_N <sub>P</sub>	strong	taste, ni	ce color, varied	vegetab	les are	Quality vs. Defect
Consumer	Product	D_1	D_2	D_2		D_N <sub>D</sub>	Instructions (facultative)
C_1	P_1	1	1	1		0	Sensory modality (visual aspect, texture, flavor, etc.)
•••		•••					Quality vs. Defect
C_N <sub>c</sub>	P_N <sub>P</sub>	1	0	1	•••	0	Quality vs. Defect
Descriptor cited or not							

Consumer	Product		D	escriptio	n		Instructions (facultative)
C_1	P_1	good 1	taste, go to	od veget o compa		ell but	Sensory modality (visual aspect, texture, flavor, etc.)
				•••			
C_N <sub>c</sub>	P_N <sub>P</sub>	strong	taste, ni	ce color, varied	vegetab	les are	Quality vs. Defect
Consumer	Product	D_1	D_2	D_2		D_N <sub>D</sub>	Instructions (facultative)
C_1	P_1	1	1	1		0	Sensory modality (visual aspect, texture, flavor, etc.)
							Quality vs. Defect
C_N <sub>C</sub>	P_N <sub>P</sub>	1	1 0 1 0				Quality vs. Defect

# **Overview**

- I. The collected data and their encoding for statistical analysis
- **II.** Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

**Rule-based data extraction** = application a particular set of rules and patterns to capture specific expressions and **transform text into data** 

<u>Rule creation</u>: Based on a **domain-specific lexicon** including **food related attributes grouped by categories** 

Rule application:

- Normalization of Free-Comment to prepare raw text for NLP
- Extraction of attributes using regular expressions (sequences of characters that specify a search pattern used by string-searching algorithms)
- Contextualization of attributes using semantic rules

<u>Rule refinement</u>: **Iterative refinement** of rules by repetitive processing to improve accuracy and performance based on **panel leader feedback** 

# **Example of Free-Comment given by a consumer**

This compote has a taste of peaches with large pieces giving it an "homemade" aspect. It is not acid but too swet and cloying.

# **Conversion of text to lower case**

this compote has a taste of peaches with large pieces giving it an "homemade" aspect. It is not acid but too swet and cloying.

# **Removal of words less than 3 characters** (except those in a predefined list)

this compote has *a* taste *of* peaches with large pieces giving *it an* "homemade" aspect. *It is* not acid but too swet and cloying.

# **Removal of special characters and multiple spaces**

*this compote has taste peaches with large pieces giving "homemade" aspect. not acid but too swet and cloying.*  **Spellchecking** using hunspell (free spell checker and morphological analyzer library, available in many programming languages)

this compote has taste peaches with large pieces giving homemade aspect. not acid but too **sweet** and cloying.

# II. Cleaning information and extracting sensory descriptors

# Search for words/expressions of interest

- attributes (related to food products)
- quantifiers (related to perceived intensities, sizes or levels of appropriateness of the attributes)
- context words (additional information on the attributes)

this <mark>compote</mark> has <mark>taste peaches</mark> with <mark>large</mark> pieces</mark> giving <mark>homemade</mark> aspect. not acid but too sweet and cloying. **Lemmatization** = replacement of words/expressions by their canonical forms (singular, masculine)

this <mark>compote</mark> has <mark>taste peaches</mark> with large pieces</mark> giving homemade aspect. not acid but too-much sweet and cloying. **Association** of quantifiers and context words with attributes (based on semantic-rules, recursive search)

this <mark>compote</mark> has <mark>taste+peach</mark> with <mark>large+piece</mark> giving homemade+aspect. not+acid but too-much+sweet and cloying.

# **Overview**

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- **III.** Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

**Objective = to "make sense" of the extracted data** and give information to the clustering algorithm **to aggregate attributes without human intervention** 

Attributes are transformed into **concepts** organized according to a **hierarchical structure** defined in the domain-specific lexicon

- Peach is a flavor attribute, belonging to the "fruity" category, and "yellow fruit" sub-category → Flavor>Fruity>Yellow-fruit>Peach (can be aggregated with "nectarine")
- Acid is a basic taste attribute, associated with the negative quantifier "not"
  → Taste>Not-acid (≠ Taste>Acid)
- Homemade is a non-categorized attribute, associated with the contextword "appearance" → NC>Homemade-appearance

# III. Semantic-based hierarchical structuration of extracted descriptors

This compote has a taste of peaches with large pieces giving it an "homemade" aspect. It is not acid but too swet and cloying.



Expression	Attribute	Modality	Hierarchy	Quantifier	Context	Concept
compote	compote	-	-	-	-	
taste peaches	peach	flavor	fruity>yellow- fruit	-	taste	flavor>fruity>yellow-fruit>peach- taste
large pieces	piece	texture	fracturable	large	-	texture>fracturable>piece-large
homemade aspect	homema de	-	-	-	aspect	NC>homemade-aspect
too swet	sweet	taste	-	too-much	-	taste>too-much-sweet
not acid	acid	taste		not	-	taste>not-acid
cloying	cloying	valence	not-pleasant	-	-	valence>not-pleasant>cloying

### **III. Semantic-based hierarchical structuration of extracted descriptors**

	А	В	c	D	E	F
1	RegularExpressionFR	AttributeFR	FullConceptFR	ProductCategory	AttributeEN	
60	solvants?	solvant	flaveur:chimique>solvant		solvent	
61	vernis	vernis	flaveur:chimique>vernis		varnish	
62	acescent?c?e?s?	acescent	saveur:acide		sour	
63	vinaigré?e?s?	vinaigre	saveur:acide		vinegar	
64	ac[ée]tiques?	acétique	flaveur:chimique>vinaigre>acétique		acetic	
65	acides? ac[ée]tiques?	acide acétique	saveur:acide		acetic acid	
56	d[ée]fauts?	défaut	flaveur:défaut		defect	
57	anomalie	anomalie	flaveur:défaut>anomalie		anomaly	
58	bouchons?	bouchon	flaveur:défaut>bouchon	vin	corked	
59	bouchonn[ée]e?s?	bouchonné	flaveur:défaut>bouchon	vin	corked	
70	probl[ée]mes?	problème	flaveur:défaut>problème		problem	
71	empyreumatiques?	empyreumatique	flaveur:empyreumatique		empyreumat	ic:
72	bouillons?	bouillon	flaveur:empyreumatique>bouillon		broth	
73	brais[ée]e?s?	braisé	flaveur:empyreumatique>braisé		braised	
74	brul[ée]e?s?	brûlé	flaveur:empyreumatique>brûlé		burnt	
75	benzol	benzol	flaveur:empyreumatique>brulé>benzol		benzol	
76	bitum[ée]e?s?	bitume	flaveur:empyreumatique>brulé>bitume		tar	
77	caoutchouc	caoutchouc	flaveur:empyreumatique>brûlé>caoutchouc		rubber	
78	crame[ée]e?s?	cramé	flaveur:empyreumatique>brûlé>cramé		burnt	
79	goudron	goudron	flaveur:empyreumatique>brulé>goudron		tar	
B0	suie	suie	flaveur:empyreumatique>brulé>suie		soot	
31	caramels?	caramel	flaveur:empyreumatique>caramel		caramel	
32	caram[ée]lis[ée]e?s?	caramélisé	flaveur:empyreumatique>caramel		caramelized	
33	confiture lait	confiture de lait	flaveur:empyreumatique>caramel>confiture-de-lait		milk jam	
34	chocolat[ée]?e?s?	chocolat	flaveur:empyreumatique>chocolat		chocolate	
35	chocolats? chauds?	chocolat chaud	flaveur:empyreumatique>chocolat>chocolat-chaud		hot chocolate	e
86	chocolats? lait	chocolat au lait	flaveur:empyreumatique>chocolat>chocolat-lait		milk chocola	te
87	chocolats? noirs?	chocolat noir	flaveur:empyreumatique>chocolat>chocolat-noir		dark chocola	te
88	cigarettes?	cigarette	flaveur:empyreumatique>fumé		cigarette	
89	fumage	fumage	flaveur:empyreumatique>fumé		smoked	
90	fum[ée]e?s?	fumé	flaveur:empyreumatique>fumé		smoked	
91	fumet	fumet	flaveur:empyreumatique>fumé		smoked	
92	bacon	bacon	flaveur:empyreumatique>fumé>bacon		bacon	
93	barbecue	barbecue	flaveur:empyreumatique>fumé>barbecue		barbecue	
94	grill[ée]?e?s?	grillé	flaveur:empyreumatique>grillé		grilled	
95	biscottes?	biscotte	flaveur:empyreumatique>grillé>biscotte		rusk	
00	pains?	pain	flaveur:empyreumatique>grillé>pain		bread	

# **Domain-specific lexicon** (food products)

# Includes generic and specific attributes

# **III. Semantic-based hierarchical structuration of extracted descriptors**

	А	В	с	D	E	F	
1	RegularExpressionFR	AttributeFR	FullConceptFR	ProductCategory	AttributeEN		
360	solvants?	solvant	flaveur:chimique>solvant		solvent		
361	vernis	vernis	flaveur:chimique>vernis		varnish		
362	acescent?c?e?s?	acescent	saveur:acide		sour		
363	vinaigré?e?s?	vinaigre	saveur:acide		vinegar		
364	ac[ée]tiques?	acétique	flaveur:chimique>vinaigre>acétique		acetic		
365	acides? ac[ée]tiques?	acide acétique	saveur:acide		acetic acid		
366	d[ée]fauts?	défaut	flaveur:défaut		defect		
367	anomalie	anomalie	flaveur:défaut>anomalie		anomaly		
368	bouchons?	bouchon	flaveur:défaut>bouchon	vin	corked		
369	bouchonn[ée]				corked		
370	probl[ée]mesî				problem		
371	empyreumatic				empyreumat	ic	
372	bouillons?	rived from	existing lexicons	and	broth		
373	brais[ée]e?s?		existing lexicons	anu	braised		
374	brul[ée]e?s?				burnt		
375	benzol	C	ontologies		benzol		
376	bitum[ée]e?s	C	into logics		tar		
377	caoutchouc				rubber		
378	crame[ée]e?sí	and undate	ed by panel leade	rs	burnt		
379	goudron	and apado	ea by parter leade		tar		
380	suie				soot		
381	caramels?	through	experimentations		caramel		
382	caram[ée]lis[é				caramelized		
383	confiture lait				milk jam		
384	chocolat[ée]?				chocolate		
385	chocolats? chauus:	chocolat chadu	naveur.empyreumatique>cnocolat>cnocolat-cnauu		hot chocolate	2	
386	chocolats? lait	chocolat au lait	flaveur:empyreumatique>chocolat>chocolat-lait		milk chocola	te	
387	chocolats? noirs?	chocolat noir	flaveur:empyreumatique>chocolat>chocolat-noir		dark chocola	te	
388	cigarettes?	cigarette	flaveur:empyreumatique>fumé		cigarette		
389	fumage	fumage	flaveur:empyreumatique>fumé		smoked		
390	fum[ée]e?s?	fumé	flaveur:empyreumatique>fumé		smoked		
391	fumet	fumet		smoked			
392	bacon	bacon		bacon			
393	barbecue	barbecue flaveur:empyreumatique>fumé>barbecue					
394	grill[ée]?e?s?	ll[ée]?e?s? grillé flaveur:empyreumatique>grillé					
395	biscottes?	biscotte	flaveur:empyreumatique>grillé>biscotte		rusk		
396	pains?	pain	flaveur:empyreumatique>grillé>pain		bread		
	concepts	exceptions stop-words	÷ · · · · · · · · · · · · · · · · · · ·		1		

**1900 attributes** organized into categories and sub-categories of attributes:

### Sensory categories: Appearance, Flavor, Taste, Texture, Trigeminal

### Affective categories: Positive/Negative valence

• Other categories: Balance, Complexity, Familiarity, Evoked context, etc.

### 160 quantifiers organized into:

- Appropriateness: Toomuch/Right/Not-enough
- Intensity: Not/Few/Very
- Size: Small/Large

# **Overview**

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- **IV.** Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

#### **Objective**:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the constraint that each element of the partition contains descriptors with the same RootConcept. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation* and Computation, 29(4), 1285–1316. https://doi.org/10.1080/03610910008813665

#### **Objective**:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the constraint that each element of the partition contains descriptors with the same RootConcept. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

	2 steps a	pproach:	
1 - Hierarchical clustering	g	2 – Pa	rtitioning (transfer) algorithm

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation* and Computation, 29(4), 1285–1316. https://doi.org/10.1080/03610910008813665

#### **Objective**:

Find a partition of the descriptors such that the resulting product discrimination is maximized under the constraint that each element of the partition contains descriptors with the same RootConcept. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

	2 steps a	pproach:	
	1 - Hierarchical clustering	<b>2</b> – Pa	artitioning (transfer) algorithm
1. 2.	Each descriptor forms a cluster Merge clusters resulting in the best increase in product discrimination among mergers that satisfy the constraint		
3.	Repeat until no merge lead to an increase in product discrimination		

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. https://doi.org/10.1080/03610910008813665 **27** 

#### **Objective**:

Find a partition of the descriptors such that the resulting product discrimination is maximized under the constraint that each element of the partition contains descriptors with the same RootConcept. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

		2 steps a	pproach:	
	1 - Hierarchical clustering	;	2 – Pa	rtitioning (transfer) algorithm
1. 2.	Each descriptor forms a cluster Merge clusters resulting in the best product discrimination among merger the constraint			
3.	Repeat until no merge lead to an product discrimination	increase in		
-	Provides a reasonable estimation of the clusters and initialization of the algorithm			

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. https://doi.org/10.1080/03610910008813665 **2** 

#### **Objective**:

Find a partition of the descriptors such that the resulting product discrimination is maximized under the constraint that each element of the partition contains descriptors with the same RootConcept. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

	2 steps ap	pproa	ch:	
1 - Hierarchical clusterin	g		<b>2 –</b> Pa	rtitioning (transfer) algorithm
Each descriptor forms a cluster Merge clusters resulting in the best product discrimination among merge the constraint Repeat until no merge lead to a product discrimination	ers that satisfy	2. Co ot dis	onsider hers an scrimina	each descriptor in turn moving it from its current cluster to every of perform the move if it increases product ation. ntil convergence
Ducuides a vecce bla activation of t	h			

→ Provides a reasonable estimation of the number of clusters and initialization of the partitioning algorithm

1.

2.

3.

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. https://doi.org/10.1080/03610910008813665 **2** 

#### **Objective**:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the constraint that each element of the partition contains descriptors with the same RootConcept. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

	2 steps a	ppr	oach:	
1 - Hierarchical clustering			2 – Partitioning (transfer) algorithm	
1. 2.	Each descriptor forms a cluster Merge clusters resulting in the best increase in product discrimination among mergers that satisfy the constraint		Consider	each descriptor in turn moving it from its current cluster to every d perform the move if it increases product ation.
3.	Repeat until no merge lead to an increase in product discrimination	3.	Repeat ur	ntil convergence
	Provides a reasonable estimation of the number of clusters and initialization of the partitioning algorithm	$\rightarrow$	Refined th	e solution from the hierarchical clustering

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation* and Computation, 29(4), 1285–1316. https://doi.org/10.1080/03610910008813665

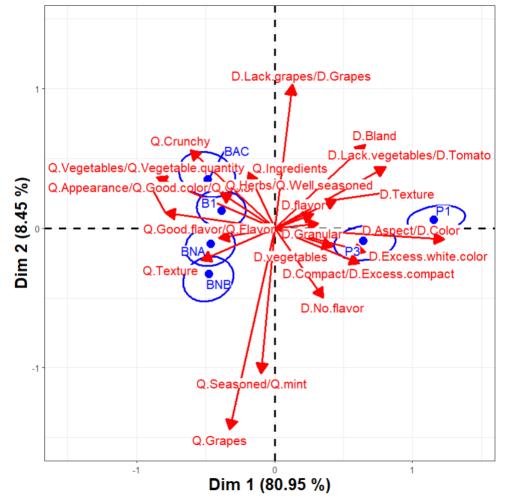
# **Overview**

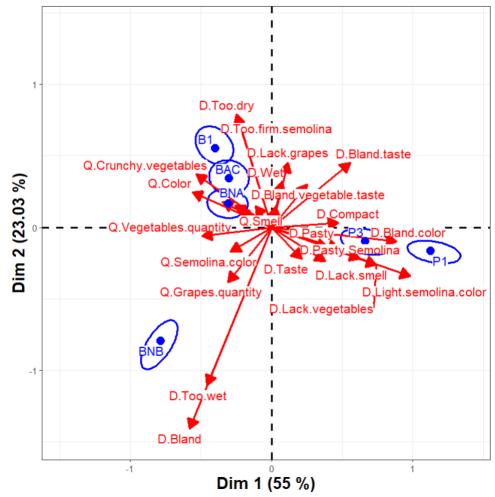
- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

#### **Automated encoding process**

#### Manual encoding process

Only descriptors having an association with at least one product are shown



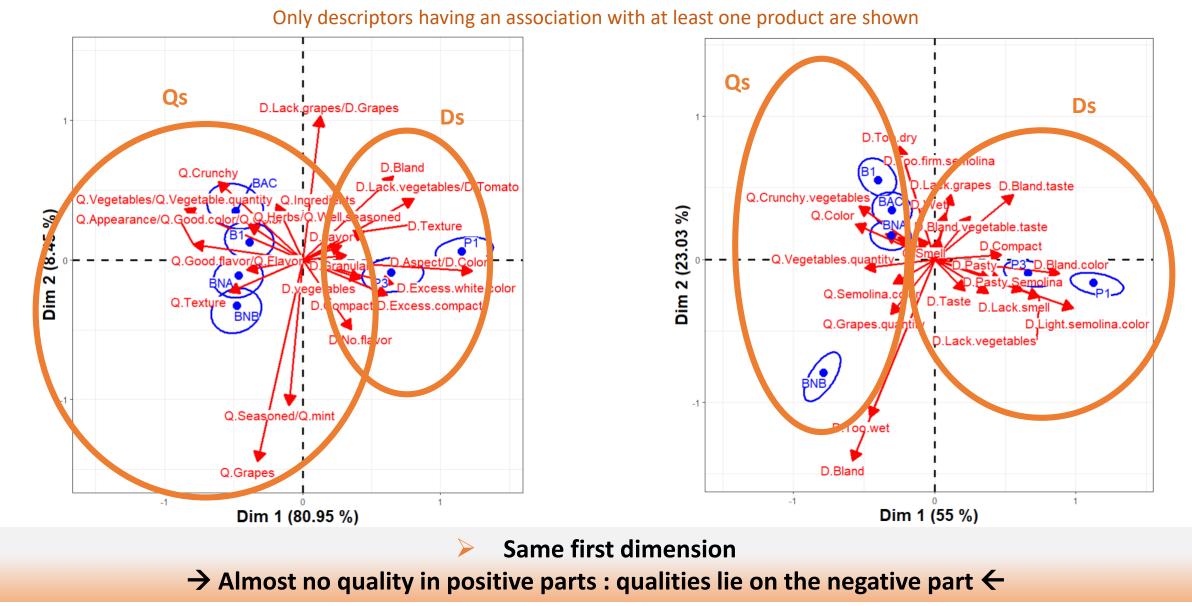


#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry .Too.firm.semolina D.Bland Q.Crunchy D.Lack.grapes\_D.Bland.taste D.Lack.vegetables/D.Tomato Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients (% (% Q.Color Q.Appearance/Q.Good exture ble.taste Dim 2 (23.03 (8.45 D.Compact -Q.Vegetables.quan D.Aspect/D.Colo .<u>Bl</u>and.colo 2 D.Excess.white.color D.vegetables Dim Q.Semolina.cgl Taste D.Lack.smell Q.Texture D.Compact/D.Excess.compact Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetables BNB Q.Seasoned/Q.mint D.Too.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %) Same first dimension

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry .Too.firm.semolina D.Bland Q.Crunchy D.Lack.grapes\_D.Bland.taste D.Lack.vegetables/D.Tomato Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients Q.Color (% (% Q.Appearance/Q.Good Dim 2 (23.03 (8.45 -Q.Vegetables.quan D.Aspect/D.Colo .Bland.colo 2 D.Excess.white.color D.vegetables Dim Q.Semolina.cgl D.Taste D.Lack.smell Q.Texture D.Compact/D.Excess.compact D.Light.semolina.color Q.Grapes.quantity D.No.flavor Ps Bs Ps Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %) Same first dimension $\rightarrow$ Product configuration $\leftarrow$

#### Automated encoding process

Manual encoding process



#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry .Too.firm.semolina D.Bland Q.Crunchy D.Lack.grapes\_D.Bland.taste D.Lack.vegetables/D.Tomato Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients (% (% Q.Color Q.Appearance/Q.Good le.taste Dim 2 (23.03 (8.45 -Q.Vegetables.quan D.Aspect/D.Color D.Bland.color 2 D.Excess.white.cold D.vegetables Dim Q.Semolina.cgl D.Taste D.Lac Q.Texture D.Compact/D.L Q.Grapes.quantity ight.semolina.color D.No.flavor D.Lack.vegetable BNB Q.Seasoned/Q.mint D.Too.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %) Same first dimension $\rightarrow$ Ps show color lightness defect $\leftarrow$

36

### Automated encoding process

(%

(8.45

2

Dim

### Manual encoding process

Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry Too.firm.semolina Q.Crunchy D.Lack.grapes D.Bland.taste D.Lack.vegetables/D.Tomato Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredien (% Q.Color Q.Appearance/Q.Good <sup>-</sup>exture /edetable.tas Dim 2 (23.03 -Q.Vegetables.quan D.Aspect/D.Color and.colo D.yegetables D.Excess.white.color Q.Semolina.cold D.Taste D.Lack.smell Q.Texture smpact/D.Excess.compact Q.Grapes.guan aht.semolina.color D.No.flavor D.Lack.vegetables BNB Q.Seasoned/Q.mint D.Too.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %)

Same first dimension

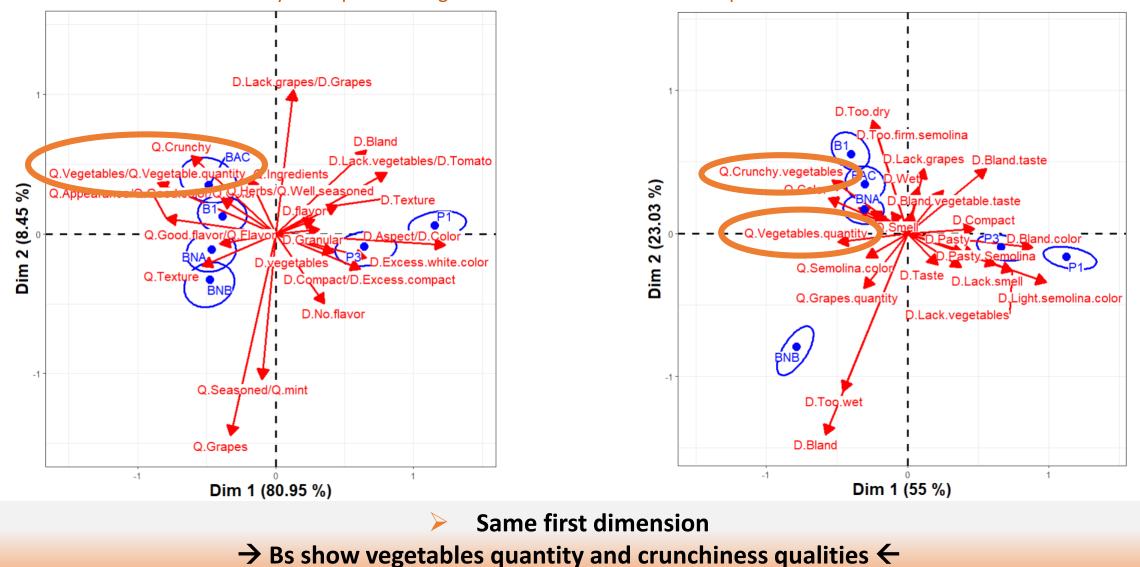
 $\rightarrow$  Ps show vegetables defects  $\leftarrow$ 

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry .Too.firm.semolina D.Bland Q.Crunchy D.Lack.grapes\_D.Bland.taste D.Lack.vegetables/D.Tomato Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients Q.Color (% (% Q.Appearance/Q.Good D.Texture Dim 2 (23.03 (8.45 D\_Compact Q.Vegetables.quant 2 D.Excess. e color Dim Q.Semolina.col Q.Texture D.Compact/D.Excess.compact D.Lack.sn Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetables BNB Q.Seasoned/Q.mint D.Too.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %) Same first dimension $\rightarrow$ Ps show compactness defects $\leftarrow$

#### Manual encoding process **Automated encoding process** Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry Too.firm.semolina D.Bland Q.Crunchy ack vege Lack.grap > D.Bland.taste oles/D.Tomato Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients Q.Crunchy.vegetable (% (% Q.Color Q.Appearance/Q.Good Dim 2 (23.03 (8.45 -Q.Vegetables.quan D.Aspect/D.Colo and.colo 2 D.Excess.white.color D.vegetables Dim Q.Semolina.cgl D.Taste D.Lack.smell Q.Texture D.Compact/D.Excess.compact Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable BNB Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %) Same first dimension $\rightarrow$ Ps show blandness defects $\leftarrow$

#### Manual encoding process **Automated encoding process** Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry Too.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato Lack.grapes D.Bland.taste etables/Q.Veget Q.Crunchy.veget antity Q.Ingredients QV (% Q.Color % Q.Appearance/Q.Good Dim 2 (23.03 (8.45 -Q.Vegetables.qua D Aspect/D Cold 2 D.vegetables .Excess.white.color Dim Q.Semolina.col Q.Texture D.Compact/D.Excess.compact D.Lack.smel D.Light.semolina.color D.No.flavor D.Lack.vegetable Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 \_1 Dim 1 (55 %) Dim 1 (80.95 %) Same first dimension $\rightarrow$ Bs show appearance/color qualities $\leftarrow$

# Automated encoding processManual encoding processOnly descriptors having an association with at least one product are shown



#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry Too.firm.semolina D.Bland Q.Crunchy Lack.grapes D.Bland.taste D.Lack.vegetables/D.Tomato Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients % (% Q.Color Q.Appearance/Q.Good.color Dim 2 (23.03 (8.45 Q.Vegetables.quan D.Aspect/D.Colo Bland.cold 2 etables .Excess.white.color Dim Q.Semolina.co Q.Texture 「aste D.Compact/D.Excess.compact D.Lack.smel Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable BNB Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %)

A additional flavor and texture quality information on the first dimension with automation

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry .Too.firm.semolina D.Bland Q.Crunchy D.Lack.grapes\_D.Bland.taste D.Lack.vegetables/D.Tomato Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients (% (% Q.Color Q.Appearance/Q.Good ole.taste Dim 2 (23.03 (8.45 -Q.Vegetables.quan D.Aspect/D.Colo .<u>Bl</u>and.colo 2 D.vegetables .Excess.white.color Dim Q.Semolina.cgl Taste D.Lack.smell Q.Texture D.Compact/D.Excess.compact Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetables BNB Q.Seasoned/Q.mint D.Too.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension**

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry Too.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato D.Lack.grapes D.Bland.taste Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients % (% Q.Color Q.Appearance/Q.Good Dim 2 (23.03 (8.45 Q.Vegetables.guan Aspect/D Cold Bland.colo 2 D.vegetables .Excess.white.color Dim Q.Semolina.cg Taste D.Lack.smel D.Compact/D.Excess.compact Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 -1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension** $\rightarrow$ BNB less "particular" with automation $\leftarrow$

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry Too.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato D.Lack.grapes D.Bland.taste Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients % (% Q.Color Q.Appearance/Q.Good Dim 2 (23.03 (8.45 Q.Vegetables.guan D Aspect/D Cold Bland.cold 2 D.Excess.white.color D.vegetables Dim Q.Semolina.cg Q.Texture Taste D.Compact/D.Excess.compact D.Lack.smel Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable Q.Seasoned/Q.mint D.Too.wet D.Bland Q.Grapes -1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension** $\rightarrow$ Lack wet defect information with automation $\leftarrow$

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry .Too.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato apes D.Bland.taste Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients Q.Color (% (% Q.Appearance/Q.Good Dim 2 (23.03 (8.45 Q.Vegetables.guan D.Aspect/D.Colo Bland.cold 2 D.vegetables .Excess.white.color Dim Q.Semolina.col Taste D.Lack.smel Q.Texture D.Compact/D.Excess.compact Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable Q.Seasoned/Q.mint D.Too.wet D.Bland Q.Grapes -1 \_1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension**

 $\rightarrow$  Lack wet defect information with automation  $\leftarrow$  BUT...

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry Too.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato D.Lack.grapes D.Bland.taste Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients % (% Q.Color Q.Appearance/Q.Good Dim 2 (23.03 (8.45 Q.Vegetables.guan D Aspect/D Cold Bland.cold 2 D.vegetables .Excess.white.color Dim Q.Semolina.cg Q.Texture D.Compact/D.Excess.compact Faste D.Lack.smel Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable BNB Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension** $\rightarrow$ Lack bland defect information with automation $\leftarrow$

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.dry .Too.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato D.Lack.grape D.Bland.taste Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients (% (% Q.Color Q.Appearance/Q.Good Dim 2 (23.03 (8.45 Q.Vegetables.guant D.Aspect/D.Colo Bland.colo 2 D.Excess.white.color D.vegetables Dim Q.Semolina.col D.Taste D.Lack.smel Q.Texture D.Compact/D.Excess.compact Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable BNB Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension**

 $\rightarrow$  Lack bland defect information with automation  $\leftarrow$  BUT...

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.drv oo.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato rapes D.Bland.taste Q.Crunchy.vegetables Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients % (% Q.Color Q.Appearance/Q.Good Dim 2 (23.03 (8.45 Q.Vegetables.guan 2 D.Excess.white.color D.vegetables Dim Q.Semolina.cg Q.Texture D.Compact/D.Excess.compact Faste D.Lack.smel Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable BNB Q.Seasoned/Q.mint D.Tog.wet D.Bland Q.Grapes -1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension**

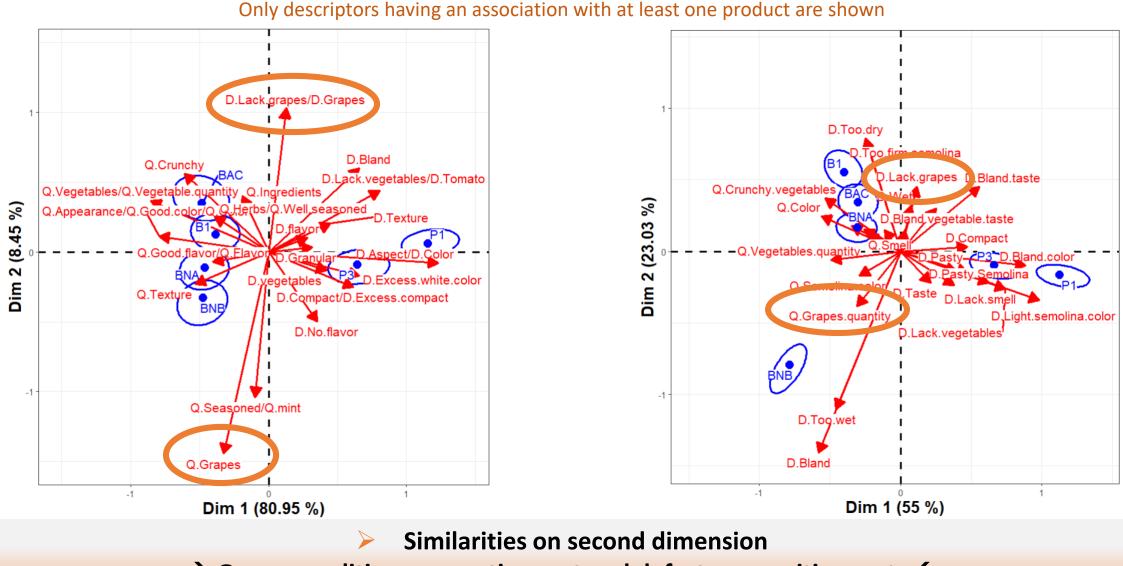
 $\rightarrow$  Lack too dry/firm defect information with automation  $\leftarrow$  (without "but")

#### **Automated encoding process** Manual encoding process Only descriptors having an association with at least one product are shown D.Lack.grapes/D.Grapes D.Too.drv Too.firm.semolina D.Bland Q.Crunchy D.Lack.vegetables/D.Tomato Lack.grapes\_D.Bland.taste Q.Crunchy.vegetable Q.Vegetables/Q.Vegetable.quantity\_Q.Ingredients Q.Color (% (% Q.Appearance/Q.Good Dim 2 (23.03 (8.45 Q.Vegetables.guan D.Aspect/D.Cold Bland.cold 2 D.Excess.white.color D.vegetables Dim Q.Semolina.cg Q.Texture D.Compact/D.Excess.compact Faste D.Lack.smel Q.Grapes.quantity D.Light.semolina.color D.No.flavor D.Lack.vegetable BNB V i Q.Seasoned/Q.min D.Too.wet D.Bland Q.Grapes -1 Dim 1 (55 %) Dim 1 (80.95 %) **Differences on second dimension**

 $\rightarrow$  Lack seasoning quality information with manual coding  $\leftarrow$  (still without "but")

### **Automated encoding process**

Manual encoding process



→ Grapes qualities on negative part and defects on positive parts ←

# **Overview**

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

# Conclusion

"Old-school" approach

- **Expert system** rather than machine learning (no need for large datasets to train AI)
- (Relatively) easy to implement
- More control on extracted data compared to pretrained models for natural language such as ChatGPT

Limitations (WIP)

- Only in French (at the moment)
- Issues with some contexts not always correctly handled
- Issues of classification with attributes having several meanings (e.g. sweet)
- Some loss of information compared to manual preprocessing

**Promising results** 

- Computation time **about 5 minutes** for a dataset (one full day for the panel leader)
- Reproducible data preprocessing
- Combination of semantics and statistics with no human intervention

Will be available as a **web app** free for non-commercial use

## **Acknowledgements**

Techni'Sens

Cécile MURSIC

Matthieu CAMILLERI

**Ronan SYMONEAUX** 

