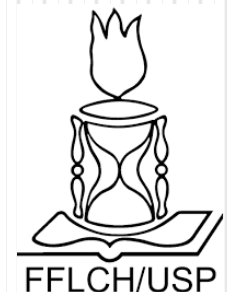


# Produire expérimentalement des corpus spécialisés pour l'analyse sémantique : méthodes et enjeux

Mariele MANCEBO-HUMBERT - UBFC/USP

Laurent GAUTIER - UBFC



# Plan de la présentation

1. Éléments de contexte
2. Méthodologie et Corpus
3. Discussion : possibilités d'exploitation des données
4. Bilan

# 1. Éléments de contexte

- Les corpus traditionnels de discours spécialisé :
  - des dictionnaires ;
  - des corpus de texte écrit : ensembles hétérogènes ;
  - des textes oraux retranscrits ;
  - des corpus multimodaux.
- Domaine viti-vinicole, filière de vins effervescents : un domaine spécialisé à dimension sensible :
  - besoin de (faire) produire des discours dans des situations spécifiques,
  - des corpus produits **expérimentalement** de façon *ad hoc*.

# 1. Éléments de contexte

- Formes d'approches traditionnelles : les « notes de dégustation » et assimilés, constitutives du :
  - Discours ambiant (spécialisé du troisième type – Petit, 2005) :
    - des guides de dégustations ;
    - des revues spécialisées ;
    - revues de presse des filières ;
    - sites web de producteurs.
  - Discours très souvent prescriptifs (Gautier, 2014), qui jouent un rôle clé en matière de recommandation = des discours maîtrisés et réfléchis :
    - Communication distanciée : ex. travail de mise en forme et de reformulation des fiches de dégustation des panels par les rédacteurs de revues et de guides.
- « [...] on ne peut donc surinterpréter le choix des termes et ignorer le fait qu'ils relèvent d'une communication distanciée [...] » (Gautier, 2018)

## 2. Méthodologie et corpus

- Deux types de corpus expérimentaux : avec et sans stimulus.

### Les corpus **sans stimulus**

- Deux objets d'étude, deux corpus :
  - le corpus *Questionnaire Crémant de Bourgogne*
  - le corpus *Questionnaire espumantes*
- consistent à faire répondre des **consommateurs** à des questions sur les produits en question, sans référent, pour une génération de paroles en dehors de toute situation de dégustation, basée sur les connaissances ou les expériences antérieures des individus, la mémoire expérientielle.
- permettent d'approcher les modes de construction sémantique des termes descripteurs en observant les stratégies de mise en discours (hésitations, reformulations, explications) et la construction des prototypes.

## 2. Méthodologie et corpus

- Approche dite « de terrain », empirico-inductive (Blanchet, 2012)
- Questionnaire unique :
  - pour les deux objets : Crémant de Bourgogne et *espumante*
  - dans les deux situations de collecte : sans et avec stimulus
- L'élaboration du questionnaire répond aux objectifs de l'étude et prévoit une analyse linguistique sémantico-cognitive (Mondada, 1998) :
  - un questionnaire semi-directif : des questions ouvertes

permettant « d'identifier les représentations cognitives, en particulier en vue de déterminer les propriétés sémantiques d'un concept [...] et la manière dont on peut identifier leurs relations à des catégories cognitives. » (Delepaut, 2009 : 164)

## 2. Méthodologie et corpus

- Questionnaire composé de 3 questions complémentaires => trois sous-corpus de parole.
- L'ordre des questions : éviter des « effets de contexte » (Delepaut, 2009 : 170) => du plus général au plus spécifique.

**Q1. Si je vous dis « Crémant de Bourgogne », à quoi pensez-vous ?**

- sous-corpus « évocation » : qui permet de verbaliser de façon spontanée et intuitive l'image du Crémant de Bourgogne présente à l'esprit des répondants.

**Q2. Si vous deviez expliquer à un ami ce qu'est un Crémant de Bourgogne, que lui diriez-vous ?**

- sous-corpus « explication » : qui permet d'accéder à un niveau d'abstraction plus élevé incitant au dépassement de l'évocation intuitive pour aborder des aspects « techniques ».

**Q3. Si vous deviez choisir trois mots ou expressions pour définir le Crémant de Bourgogne, lesquels choisiriez-vous ?**

- sous-corpus « définition » : qui permet de verbaliser des traits définitoires à partir desquels les répondants conceptualisent l'objet Crémant de Bourgogne.

## 2. Méthodologie et corpus

- Corpus *Questionnaire Crémant de Bourgogne*
  - Distribué par une agence de sondage : **2250** répondants.
- Corpus *Questionnaire espumante*
  - Distribué en ligne : **528** répondants.
- Critères sociogéographiques genre, âge et région équilibrés pour le Crémant.
- Catégories de consommation => accès aux questions ouvertes :
  - ✓ vin ;
  - ✓ vin pétillant, mousseux ;
  - ✓ champagne ;
  - ✓ spiritueux et alcools forts.



## 2. Méthodologie et corpus

- Corpus *Questionnaire Crémant de Bourgogne* et *Questionnaire Espumantes*
- Données statistiques de base (AntConc) :

	Sous-corpus	Types	Hapax	Tokens
<i>Questionnaire CB</i>	• <i>évocation</i>	1 156	-	14 028
	• <i>explication</i>	1 330	-	19 996
	• <i>définition</i>	849	-	8 117
	<b>TOTAL <i>Questionnaire CB</i></b>	<b>2 033</b>	<b>951</b>	<b>42 141</b>
	Sous-corpus	Types	Hapax	Tokens
<i>Questionnaire espumantes</i>	• <i>évocation</i>	864	-	3 649
	• <i>explication</i>	939	-	5 747
	• <i>définition</i>	514	-	2 160
	<b>TOTAL <i>Questionnaire espumantes</i></b>	<b>1 560</b>	<b>835</b>	<b>11 556</b>

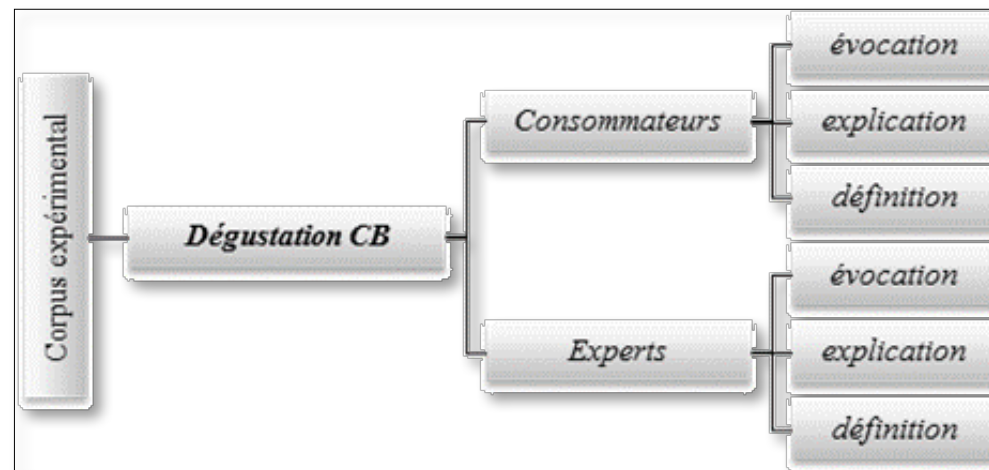
## 2. Méthodologie et corpus

### Le corpus **avec stimulus**

- Le corpus *Dégustation Crémant de Bourgogne*
  - Conçu en collaboration avec AgroSup Dijon et l'UPECB :
    - montée en gamme du Crémant de Bourgogne ;
    - évènement inédit intitulé « Les Éminents de Bourgogne » ;
    - une séance de dégustation originale réunissant *consommateurs* et *experts*.
  - Il s'agit ici de faire répondre des **consommateurs** et des **experts** à des questions permettant de collecter un corpus à partir de la génération de paroles en situation de dégustation, où la mise en mots est couplée et déclenchée par le sensoriel.

## 2. Méthodologie et corpus

- Mêmes trois questions => même trois sous-corpus de parole
  - 2 panels : expert et consommateurs



- Evènement centré sur une analyse sensorielle :
  - nécessité de collecter des données nombreuses ;
  - suivre un protocole de dégustation précis, basé sur des plans d'expériences afin de garantir la validité des résultats.

## 2. Méthodologie et corpus

- L'enchaînement des étapes, la réalisation de l'expérience de dégustations et les analyses de résultats ont alors été organisés de façon à respecter une triple exigence :
  1. chaque dégustateur évalue un même nombre de vins ;
  2. chaque vin est évalué un même nombre de fois ;
  3. toute paire de vins est évaluée un même nombre de fois ; afin d'éviter des facteurs comme « le rang d'évaluation du vin, la succession des vins ou le dégustateur » lors de la mesure sensorielle.

## 2. Méthodologie et corpus

- Trois étapes :
    - deux étapes d'analyse sensorielle
    - une étape de constitution de données textuelles expérimentales (réponses au questionnaire)
1. Test hédonique à l'aveugle = j'aime ou je n'aime pas
  2. Test bouteille et dégustation (conso) et Test par attributs (pro)
  3. Mise en mots du Crémant de Bourgogne
    - collecte de données à deux reprises : 2016 et 2017 à l'Auditorium de Dijon

## 2. Méthodologie et corpus

- Recrutement de panels :
- Consommateurs : plateforme ChemoSens du CSGA, critères :
  - lieu de résidence : Dijon, agglomération et proche alentours ;
  - consommation d'effervescents : consommateurs réguliers ou occasionnels ;
  - genre : équilibre homme/femme ;
  - tranches d'âge : équilibre entre trois tranches d'âge :
    - entre 18 et 39 ans, entre 40 et 59 ans, et 60 ans et plus
- **104** répondants pour la première édition et **95** pour la deuxième = **199 consommateurs.**
- Experts : à la charge de l'UPECB et il compte :
  - des opérateurs de la filière ;
  - des œnologues ;
  - des cavistes ;
  - des commerciaux ;
  - des prescripteurs au sens large, comme la presse spécialisée ou des enseignants du domaine à condition d'être reconnu comme un professionnel par la filière.
- **54** répondants pour la première édition et **70** pour la deuxième = **124 professionnels.**

## 2. Méthodologie et corpus

- Corpus *dégustation Crémant de Bourgogne*
- Données statistiques de base (AntConc) :

	Sous-corpus	Types	Hapax	Tokens
<i>Dégustation CB</i>	<b>Consommateurs</b>	1 212	656	<b>8 286</b>
	• <i>évocation</i>	693	-	3 070
	• <i>explication</i>	735	-	3 948
	• <i>définition</i>	420	-	1 268
	<b>Professionnels</b>	975	-	<b>5 247</b>
	• <i>évocation</i>	509	-	1 916
	• <i>explication</i>	617	-	2 666
	• <i>définition</i>	252	-	665
	<b>TOTAL Dégustation CB</b>	<b>1 684</b>	<b>915</b>	<b>13 533</b>

# 3. Discussion

- Une fois les corpus collectés, il existe des nombreuses possibilités de traitements et analyses :
  - cette étude se base sur la linguistique « de corpus » (Tognini-Bonelli, 2001)
- Méthodologie inductive :
  - corpus = le point de départ
  - courant *corpus-driven* ou *bottom-up*

l'analyse quantitative qui intervient en premier lieu, suivie par la description qualitative des fonctions discursives dans les textes segmentés à partir des critères linguistiques qualifiables (Biber *et al.*, 2007).



# 3. Discussion

- TXM : les questions de fréquence, la combinatoire
- Variables corpus *dégustation CB et questionnaire CB* :
- **propriété** : {année}, {type-corpus}, {statut}, {sexe}, {recode-age}, {conso}
- **structure** : {question}

```
<?xml version="1.0" encoding="UTF-8"?>
<text id="1" annee="2016" type-corpus="avec stimulus" statut="professionnel" sexe="homme"
recode-age="60 ans ou plus" conso="NR" profession="commercial">
<body>
<question n="1">Effervescence, fraîcheur, terroir, tradition française.
</question><question n="2">Un vin effervescent élaboré suivant la méthode traditionnelle
à partir de vin d'appellation contrôlé. Des vins d'un grand terroir, la Bourgogne, le
seul vin qui assemble les 2 plus grands cépages : le Pinot Noir et le Chardonnay.
</question><question n="3">Elégance, fraîcheur, authenticité.</question>
</body>
</text>
```


```
<?xml version="1.0" encoding="UTF-8"?>
<text id="1" annee="2016" type-corpus="sans stimulus" statut="consommateur" sexe="homme"
recode-age="entre 40 et 59 ans" conso="entre 3 et 6 fois par an" region="Paris,
Ile-de-France">
<body>
<question n="1">un champagne bas de gamme</question><question n="2">vin pétillant pour
cocktail</question><question n="3">pétillant régional cocktail</question>
</body>
</text>
```

# 3. Discussion

- De nombreuses requêtes sont possibles :
  - fréquence de mots ;
  - fréquence de mots lemmatisés ;
  - requête par propriété : NOM, ADJ, VER.\*, ADV... ;
  - cooccurrence / combinatoire ;
  - créer des partitions ou des sous-corpus ;
  - réorganiser le lexique dans la table de matière ;
  - calcul des spécifiés ;
  - des Analyses Factorielles de Correspondance (AFC), etc.

# 3. Discussion

- Ex. Corpus dégustation CB :
- *Bulle(s)* : cooccurrent

Requête :  [frlemma="bulle"]

Propriétés des cooccurrents : word  Seuils : Fmin ≥ 2 Cmin ≥ 2 Indice ≥ 2

Contexte :  forme  structure   Contexte gauche actif  Contexte droit actif  inclure la structure contenant le pivot

de - 2 à - 0 et de 0 à 2

Cooccurrent	Fréquence	Cofréquence	Indice	Distance moyenne
,	1385	139	22	,8
.	1162	72	3	,8
un	516	11	-2	1,7
des	201	32	10	,1
finesse	37	11	6	1,1
fines	8	8	11	,0
plein	5	3	3	,0
fine	5	4	4	,2
ouvrir	3	2	2	1,5
charme	3	3	4	1,3
élégantes	2	2	2	1,0

# 3. Discussion

- Ex. Corpus dégustation CB :
- *Finesse* : concordance

Requête:  Pivot: word

Clés de tri: #1  #2  #3  #4

1 - 11 / 11

text_id	Contexte gauche	Pivot	Contexte droit
3-corpus-degustation-pourxml_000002_min	un vin pétillant dont le goût, la	finesse de la bulle	pourrait tromper les amateurs de champagne tellement c'est proche. pétillant
3-corpus-degustation-pourxml_000003_min	avec un élevage spécifique issu des cépages bourguignons	bulles, finesse	, aromatique
3-corpus-degustation-pourxml_000006_min	fête,	bulles, finesse	, un vin effervescent du champagne fait en bourgogne bulles, vin
3-corpus-degustation-pourxml_000020_min	à bulle au caractère noble mais avec une	finesse de bulle	. crémant = noble, minéral, fin.
3-corpus-degustation-pourxml_000046_min	plus ou moins sec fraîcheur et netteté,	finesse des bulles	, persistance dégageant des bulles
3-corpus-degustation-pourxml_000085_min	champagnes dans une autre aire d'appellation.	finesse, bulles	, arômes
3-corpus-degustation-pourxml_000114_min	, sa qualité, son prestige, ses	bulles tout en finesse	en font un véritable délice. sobriété, diversité, excellence.
3-corpus-degustation-pourxml_000232_min	champagne de terroir, la bourgogne avec des	bulles finesse	, tradition, qualité du travail
3-corpus-degustation-pourxml_000253_min	se suppléer à un champagne à l'apéritif	bulle, finesse	, élégance
3-corpus-degustation-pourxml_000259_min	le crémant m'évoque l'élégance, la	finesse des bulles	sur des terroirs riches bourguignons. c'est se faire plaisir avec
3-corpus-degustation-pourxml_000284_min	et parfois (ou souvent) bien meilleur	bulles, finesse	, légèreté.

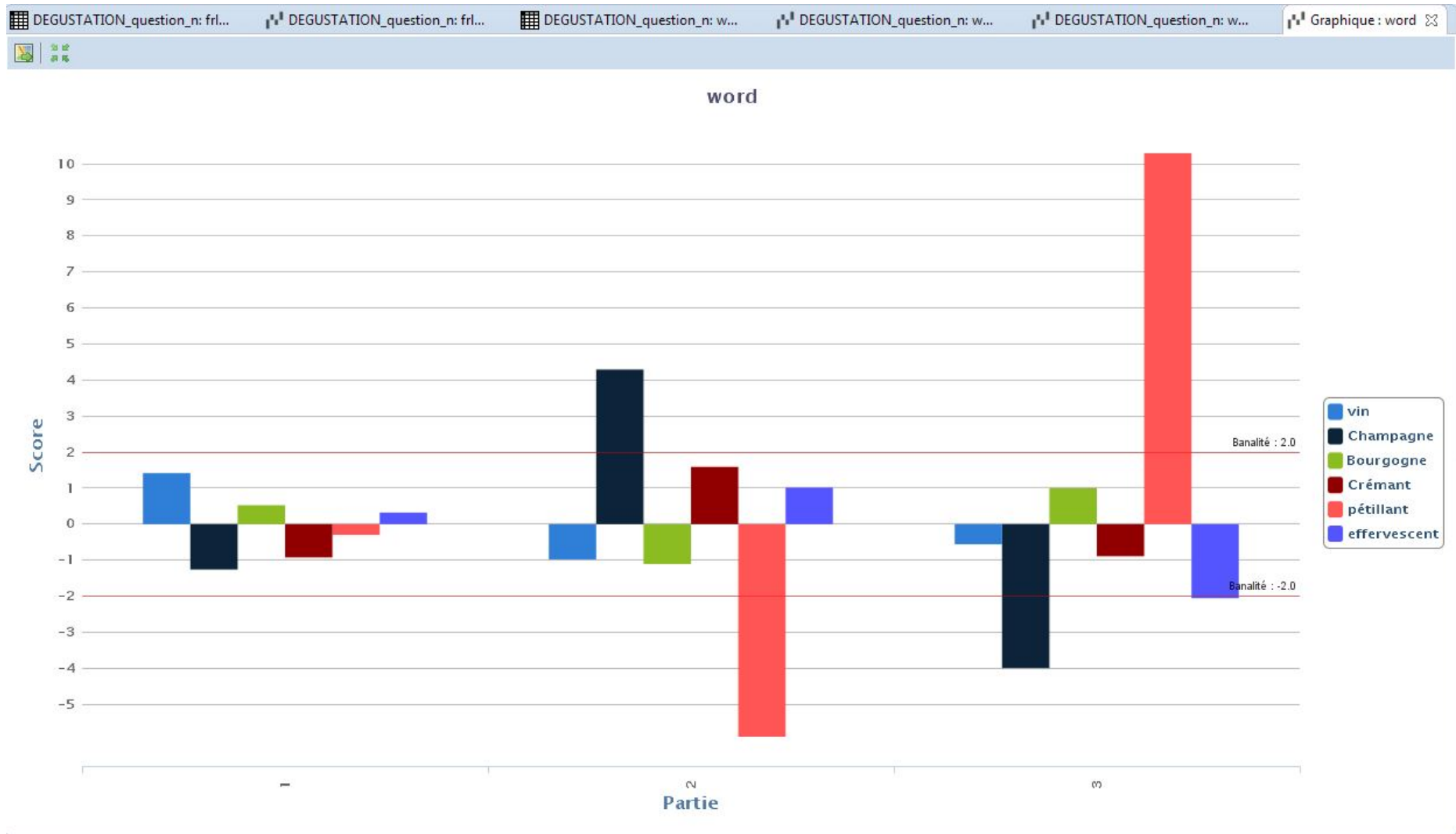
# 3. Discussion

- Ex. Corpus dégustation CB :
- Table lexicale = par question = lexique désignation et classification du produit

Unités	Fréquence T 1415	1 t=508	score	2 t=736	score	3 t=171	score
vin	404	160	1,4	199	-1,0	45	-0,6
Champagne	287	91	-1,3	179	4,3	17	-4,0
Bourgogne	275	103	0,5	132	-1,1	40	1,0
Crémant	174	55	-0,9	103	1,6	16	-0,9
pétillant	149	53	-0,3	50	-5,9	46	10,3
effervescent	126	46	0,3	73	1,0	7	-2,1

# 3. Discussion

- Table lexicale = par question = lexique désignation et classification du produit



## 4. Bilan

- Re-examen nécessaire de la place de la « terminologie » (wüstérienne) dans la caractérisation du discours professionnel visé :
  - Opposition système / usage => en discours professionnel, la terminologie ne peut être que *usage based*
  - Terminologie wüstérienne est saisie dans les textes réglementaires (cahiers des charges)
  - La force de la donnée située réside dans la mise en relation (*linking*) de phénomènes de saillance contextuelle tout autant constitutifs de l'emploi *in vivo* des termes que leurs définitions.
- Production de données expérimentales directement liée à l'objet, indispensable pour tout ce qui a une dimension sensible tributaires de communautés discursives différentes.

# Merci de votre attention !

Mariele MANCEBO-HUMBERT

[mariele.mancebo-garcia@u-bourgogne.fr](mailto:mariele.mancebo-garcia@u-bourgogne.fr)

Laurent GAUTIER

[laurent.gautier@u-bourgogne.fr](mailto:laurent.gautier@u-bourgogne.fr)

