HAL
open science

The gastronomic meal of the French through the tweets of Michelin star-rated chefs: characterization of the cultural heritage, and extraction of techniques and professional gestures

Julien Longhi, Zakarya Després, Claudia Marinica, Marcilhac Vincent, Felipe Diaz-Marin

HAL Id: hal-02996735

https://hal.science/hal-02996735

Submitted on 23 Apr 2024

The gastronomic meal of the French through the tweets of Michelin star-rated chefs: characterization of the cultural heritage, and extraction of techniques and professional gestures

Julien Longhi* **, Zakarya Després**, Claudia Marinica** ***, Vincent Marcilhac* ****, Felipe Diaz Marin****

*AGORA, EA7392, Université Paris-Seine
**Institut des Humanités Numériques de l'Université de Cergy-Pontoise
***ETIS UMR 8051, Université Paris-Seine, Université de Cergy-Pontoise, ENSEA, CNRS
*****Pôle de Gastronomie de l'Université de Cergy-Pontoise
{Julien.Longhi, Zakarya.Després, Claudia.Marinica, Vincent.Marcilhac, Felipe.Diaz-Marin}@u-cergy.fr

Abstract

This paper presents how the use of digital corpora – sourced from Twitter and Instagram accounts pertaining to two and three Michelin Star-rated Chefs respectively - may help reconstruct cooking techniques as well as professional gestures. Our approach, articulating data sciences and semiological analysis, proposes to take advantage of the heritage dimension embedded in the aforementioned social media. These social media are being used by two and three Michelin-star rated Chefs to convey knowledge and practices; these particular data are perceived as extremely useful for gastronomy experts to make sense of and make use of. We focus in particular on the most favoured culinary techniques, and on how these techniques transmit sensations and perceptions. From a methodological viewpoint, we narrow down on interdependencies between textual data and images/videos that illustrate techniques and gestures. For example, verbs contained in the text corpus shall enable extracting culinary techniques (i.e. *stuffing*, *puffing*, *infuse*) but also ways of experiencing gastronomy as such (i.e. *discover*, *taste*), which shows in turn that techniques are also related to the experience they provide. In addition, some adjectives may show qualities attributed through the preparation process, (i.e.: *confit*, *crispy*, *greasy*). The use of these adjectives demonstrates that the sensory description of the dishes stands at the core of Chefs' culinary rhetoric. Generally speaking, despite the brief format of messages, this research shows the interest of such an analytical method to extract and represent culinary techniques of the great Chefs.

Keywords: tweets, chefs, heritage, perception, gastronomy, textometry

1. Introduction

Since 2010, the gastronomic meal of the French has been inscribed on the representative list of the intangible cultural heritage of humanity: according to the definition from UNESCO[1], its important components are "the careful selection of dishes from a constantly growing repertoire of recipes; the purchase of good, preferably local products whose flavours go well together; the pairing of food with wine; the setting of a beautiful table; and specific actions during consumption, such as smelling and tasting items at the table." When we treat the gastronomic meal of the French as intangible heritage, we are not only interested in the dishes that are part of the French gastronomy, but also in the people who allow the preservation and the transmission of this heritage which, in the case material works of art, can be compared to curators or restorers. The Ministry of Agriculture also specifies that it is a heritage to be transmitted, and that it is important to safeguard the gourmet meal of the French. The paper is organized as follows: Section 2 presents the data we have processed. Section 3 presents the methodology and Section 4 presents the preliminary results. The last section concludes the article.

2. Building corpora

Among this list of restaurants awarded by the Michelin Guide, we selected those who had a Twitter account or whose chef manages his own Twitter account, or both. We extracted 47,923 tweets from these 61 accounts, which represents 775,754 words. We used a Python script based on the Tweepy library, a wrapper for the Twitter API that collects tweets from a given account. Some of those accounts are much more prolific than others: 7% of the accounts represent 25% of our corpus. We have chosen to focus on chefs who have obtained two or three stars in the Michelin Guide: the star attribution is based on identical criteria in order to guarantee the coherence of the selection. These criteria are five in number: quality of products, mastery of cooking and flavors, personality of the chef in his cuisine, quality / price ratio and regularity over time and over the entire card. The stars only evaluate "what is in the plate"; they only reward the quality of the cuisine. Three stars indicate that it is "a remarkable cuisine, worth the trip" and two stars means "an excellent table worth a detour". The choice of the Michelin Guide and the chefs rewarded by two or three stars as a sample of our study is justified by the fact that the Michelin Guide remains in France the reference gastronomic guide whose classification is authoritative and by the fact that the attribution of two or three stars brings great notoriety to restaurants and distinguished chefs, whose speech is relayed by the media.

3. Methodology

Textometry is an instrumented approach to corpus analysis, articulating quantitative syntheses and analyzes including text (Lebart & Salem, 1994). Textometry implements differential principles. The approach highlights similarities and differences observed in the corpus according to the representation dimensions considered (lexical, grammatical, phonetic, or prosodic ones, etc). Textometry

---

[1] https://ich.unesco.org/en/RL/gastronomic-meal-of-the-french-00437

establishes contextual and contrastive modeling (Pincemin 2012) and is particularly relevant to corpus exploitation in human and social sciences: detailed and global observation of different texts while remaining close to them, and highlights the fact that language is an important observation field for human and social sciences. The Iramuteq[2] software offers a set of analysis procedures for the description of a textual corpus.

One of its principal methods is Alceste. It segments a corpus into context units, to make comparisons and groupings of the segmented corpus according to the lexemes contained within it, and then to seek stable distributions (Reinert, 1998). The choice of Iramuteq for this exploratory work is motivated by our double conceptual interest for the links between forms and themes on the one hand, and between forms and profiling of another side. This is part of the Theory of Discursive Objects (Longhi, 2015), which is based on the concept of discourse object (Longhi, 2008), the Theory of Semantic Forms (Cadiot and Visetti, 2001), which mobilizes the concepts of motifs, profiles and themes. Also, we retain the lexical classification (Ratinaud and Marchand, 2015) implemented in Iramuteq because it allows to bring out the specific themes of the leaders is to group lexical worlds and highlight the general themes of the corpus, the method seeking to "give an account of the internal order of a speech, to highlight its lexical worlds"[3]. It follows from the factorial analysis of correspondence. In our case the corpus is of reasonable size, and the software offers three major groups, as shown in Figure 1:

class 1	class 2	class 3
merci	monnaiedeparis	maurocolagreco
cuisine	flaveur	thank
grand	pomme	new
beau	tomate	best
équipe	fleur	great

Figure 1: Classification according to Iramuteq (see the complete classification in appendix)

We find three classes, or lexical worlds, in this corpus. A large part of the corpus, represented by the first class with almost 60% of the terms, is related to promotion, and another part, represented by the third class and by 24% of the terms, is also linked to promotion, but in English.
An example of tweets of this type is given in Figure 2.

Figure 2: romotion by Jean-François Piège.

However, almost 16% (second class) are messages containing actual culinary information, with terms relating to ingredients or techniques / recipes.

By filtering the corpus to form a sub-corpus specific to this class, we can have a less noisy vision of what could be the culinary heritage through the digital speeches of the great chefs. For this, we use the function Characteristics Text Segments (ST) of the relevant class and then export all ST in the class, to form, with their reunification, a new corpus.

4. Preliminary results

The analysis of this new corpus then makes it possible to use other functionalities, such as the similarity analysis, which produces graphs from the R igraph library. The input table is an attendance / absence table. The similarity matrix is calculated from one of the suggested indices. Most of the indices offered come from the R proxy library:
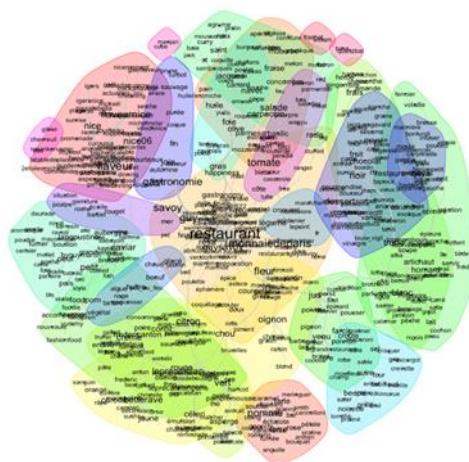
Figure 3: Similarity analysis (larger picture in Appendix)

We have a set of clusters that can find ingredients and their associations, parts of revenue perceptible through designations of dishes, or assortments or specifications.
An interesting way to experiment with this is to use the part-of-speech tagging function to look at the grammatical category of words used in the corpus, and the most common in each category (Figure 4).

---

[2] cf http://www.iramuteq.org

[3] cf https://datahist.hypotheses.org/11

verbs		nouns		adjectives	
stuff	137	tomato	349	black	261
laugh	98	flower	329	green	254
cook	97	gastronomy	321	small	223
smoke	89	apple	320	roasted	192
blow	72	lemon	270	fresh	190

Figure 4 : Top 5 most frequent words by grammatical categories, with their number of occurrences in the corpus (complete list in appendix)

We note that with the verbs, can be extracted culinary techniques (*stuff*, *blow*, *brew*) but also ways of living the dining experience (*discover*, *taste*) which confirms our idea that these data have a heritage interest. For example, the tweets in Figure 5 present dishes using techniques and products that fit into the heritage process presented.

Figure 5 : Tweets linked to culinary techniques

The category of nouns is interesting from a heritage point of view as well, because despite the corpus of two and three star chefs, the ingredients mentioned are not necessarily luxurious and we can thus think that the French culinary heritage is based on certain traditional products. terroir (*apple*, *onion*, *cabbage*). This link between heritage and terroir will have to be deepened. Moreover, it would be interesting to compare these results with the work carried out in the social sciences (notably the sociologist Claude Fischler) on the ingredients mentioned in the specialties mentioned by the three Michelin-starred chefs in the Michelin Guide: truffles, lobster, caviar or chocolate are among the most mentioned ingredients; but it is observed that more and more vegetables are mentioned in the specialties, which testifies to a new status of these products in the haute-cuisine. Finally, many adjectives of color are mentioned, indicating that the names used are often specified, either because the ingredient itself contains a color adjective (*green bean*), or because it is a specificity of the recipe. Other adjectives concern more the patrimonial dimension, through qualities attributed via the preparation (*confit*, *crispy*, *greasy*). The use of these adjectives shows that the sensory description of the dishes is the heart of chefs' culinary rhetoric. A more specific work of associations between these categories remains to be done, in order to grasp the combinatorial complexity of these parts of speech.

5. Conclusions

In this paper we present a work in progress with preliminary results around the French gastronomic heritage. To this end, we analyzed the tweets sent by chefs or two or three stars restaurants to the Michelin Guide. In the continuation of the works, an extension of the corpus to the one star chefs will be possible. It will nevertheless be necessary to check the consequences of such an extension in quantitative terms (number of accounts and quantity of data) and qualitative (specificities of this corpus and relevance for the general problem). Finally, an educational dimension can be seen, for cooking schools, since some tweets even give access to cooking recipes explanation videos. The scope of the analysis is therefore multiple: heritage, techniques, pedagogy, and transmission.

6. References

Cadiot P., Visetti Y.-M. (2001). *Pour une théorie des formes sémantiques* : PUF.

Lebart L., Salem A. (1994). *Statistique textuelle* : Dunod.

Longhi J. (2008). *Objets discursifs et doxa. Essai de sémantique discursive*. Paris : l'Harmattan.

Longhi J. (2015). *La Théorie des objets discursifs : concepts, méthodes, contributions*. Mémoire d'HDR : Université de Cergy-Pontoise.

Pincemin B. (2011). *Sémantique interprétative et textométrie – Version abrégée*, Corpus, 10 , pp. 259--269.

Ratinaud P., Marchand P. (2015). Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014), *Mots – Les langages du politique*, 108, pp. 57--77.

Reinert, M. (1998). Quel objet pour une analyse statistique du discours? Quelques réflexions à propos de la réponse Alceste. *Actes des 4èmes JADT*. URL : http://lexicometrica.univ-paris3.fr/jadt/jadt1998/reinert.htm

APPENDIX

classe 2

15.9 %

classe 1

59.7 %

classe 3

24.4 %

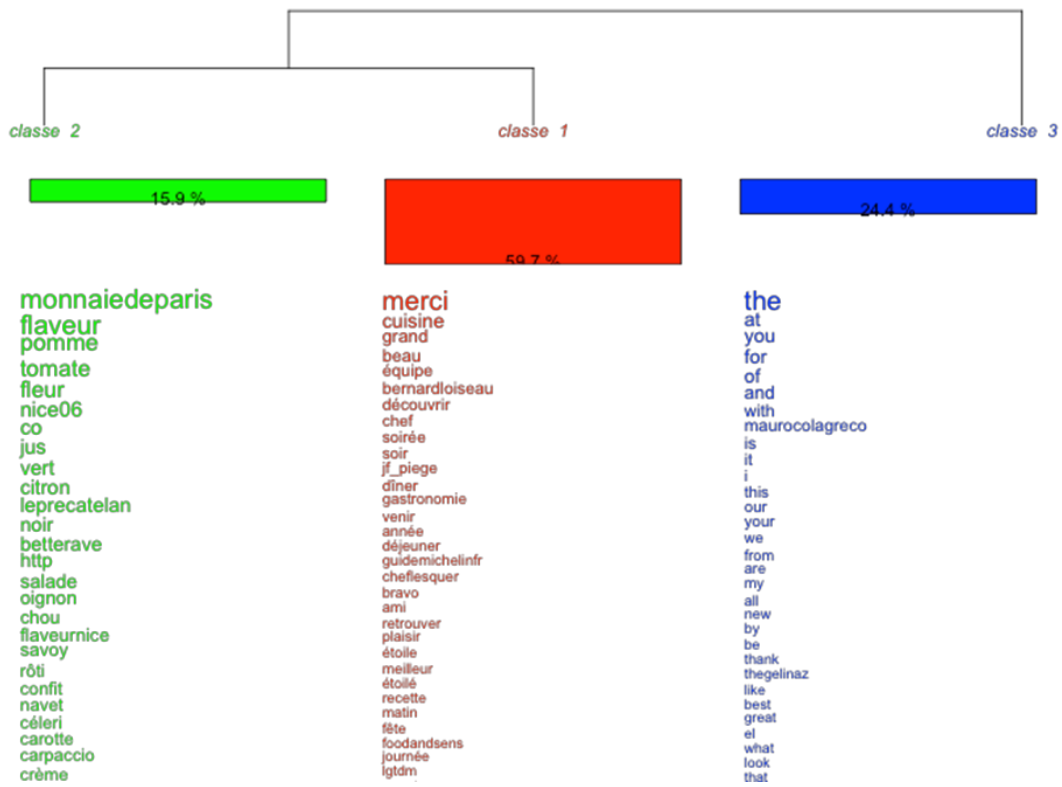| classe 2	classe 1	classe 3
monnaiedeparis	merci	the
flaveur	cuisine	at
pomme	grand	you
tomate	beau	for
fleur	équipe	of
nice06	bernardloiseau	and
co	découvrir	with
jus	chef	maurocolagreco
vert	soirée	is
citron	soir	it
leprecatelan	jf_piege	i
noir	dîner	this
betterave	gastronomie	our
http	venir	your
salade	année	we
oignon	déjeuner	from
chou	guidemichelinfr	are
flaveurnice	cheflesquer	my
savoy	bravo	all
rôti	ami	new
confit	retrouver	by
navet	plaisir	be
céleri	étoile	thank
carotte	meilleur	thegelinaz
carpaccio	étoilé	like
crème	recette	best
	matin	great
	fête	el
	foodandsens	what
	journée	look
	lgtdm	that

Figure 1: Complete classification according to Iramuteq
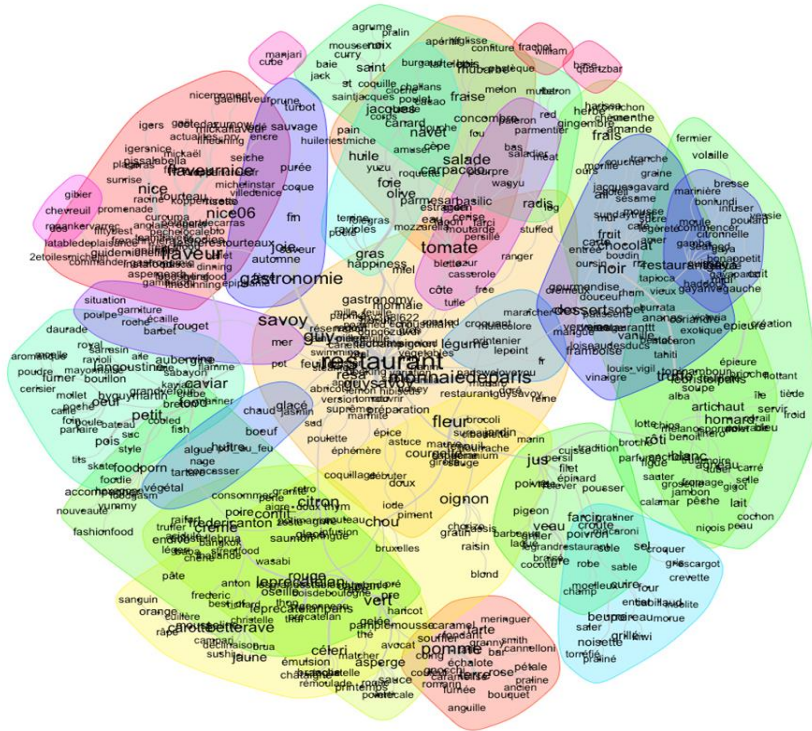
Figure 3: Similarity analysis

Forme	Freq.	Types	Forme	Freq.	Types	Forme	Freq.	Types
farcir	137	ver	tomate	349	nom	noir	261	adj
rire	98	ver	fleur	329	nom	vert	254	adj
cuire	97	ver	gastronomie	321	nom	petit	223	adj
fumer	89	ver	pomme	320	nom	rôti	192	adj
souffler	72	ver	citron	270	nom	frais	190	adj
accompagner	67	ver	jus	244	nom	confit	182	adj
griller	65	ver	truffe	242	nom	blanc	173	adj
découvrir	61	ver	nice	238	nom	gras	162	adj
servir	55	ver	caviar	236	nom	rouge	160	adj
consommer	49	ver	oignon	232	nom	jaune	138	adj
commencer	41	ver	salade	227	nom	nouveau	117	adj
saler	39	ver	dessert	227	nom	grillé	100	adj
pousser	39	ver	betterave	225	nom	glacé	96	adj
déguster	34	ver	chou	215	nom	cuit	77	adj
infuser	30	ver	crème	202	nom	croustillant	67	adj
mettre	29	ver	joie	198	nom	sauvage	65	adj
croquer	28	ver	terre	189	nom	bleu	63	adj
parfaire	25	ver	carpaccio	182	nom	premier	62	adj
concasser	25	ver	céleri	177	nom	doux	62	adj
retrouver	22	ver	homard	175	nom	beau	60	adj
matcher	21	ver	navet	174	nom	végétal	56	adj
goûter	21	ver	carotte	173	nom	gourmand	51	adj
cuisiner	21	ver	veau	169	nom	tartare	50	adj
truffer	20	ver	foie	164	nom	nouvelle	48	adj
sauter	20	ver	chocolat	163	nom	chaud	45	adj
régaler	19	ver	fraise	162	nom	pourpre	44	adj
			légume	158	nom	sablé	43	adj
			ail	155	nom	mini	43	adj
			oeuf	151	nom	feuilleté	43	adj
			tarte	149	nom	suprême	41	adj
			huile	148	nom	gris	41	adj
						mauve	40	adj
						velouté	39	adj
						poché	38	adj

Figure 4 : Most frequent words by grammatical categories, with their number of occurrences in the corpus