

*Et tout (ça), nanana, et caetera...*  
**Conditions théoriques et méthodologiques pour l'observation des *restituteurs d'ensemble par inférence***

Emmanuelle Guerin

Université Sorbonne nouvelle

[emmanuelle.guerin@sorbonne-nouvelle.fr](mailto:emmanuelle.guerin@sorbonne-nouvelle.fr)

<https://orcid.org/0000-0003-3643-5762>

Reçu le 13/10/2019, accepté le 4/3/2020, publié le 5/11/2020 selon les termes de la licence  
*Creative Commons Attribution 4.0 International (CC BY 4.0)*

**Résumé** : Dans cette contribution, on cherche à illustrer une approche de l'analyse de la langue qui pose la nécessité de recourir aux informations caractérisant une situation de communication pour comprendre le fonctionnement des unités linguistiques. Il s'agit ici de montrer en quoi les données relevant de ce que la tradition saussurienne rangerait du côté de la 'linguistique externe' peuvent éclairer, voire être indispensables, pour décrire le fonctionnement de certaines unités, parmi lesquelles ce que je nomme les *restituteurs d'ensemble par inférence* (REPI) comme *et tout (ça), nanana, et caetera...* Compte tenu de cette perspective, il s'agit d'interroger la nature des données mises à disposition, leur accessibilité, dans les corpus *OFROM* et *MPF*. Autrement dit, on se demande dans quelle mesure les cadres théoriques et méthodologiques qui articulent la constitution de ces deux corpus s'accordent avec une approche de la langue qui place les locuteurs en interaction à l'initiale de l'analyse.

**Abstract**: In this paper, I illustrate an approach to linguistic analysis that uses information about the communicative situation to understand the functioning of linguistic units. The aim is to show how data that would be classified as language-external in the Saussurian tradition can aid in or even be indispensable to the functional analysis of certain units. This is particularly true, as I argue, for what I call *restituteurs d'ensemble par inférence* (REPI) like *et tout (ça), nanana, and et caetera...* Specifically, this paper investigates to what extent the theoretical and methodological frameworks underlying the design of the *OFROM* and *MPF* corpora agree with an approach to language that places speakers in interaction at the initial stage of the analysis.

## 1 Introduction

[1] Cette contribution s'intéresse à une série d'unités, telles que *et tout (ça)*, *nanana*, *et caetera*, fréquemment observables dans les discours ordinaires. Si cette série n'est pas homogène sur le plan de la nature de ses unités, sa pertinence tient au fait que lesdites unités semblent avoir des caractéristiques fonctionnelles communes au niveau du discours en interaction. Partant, il est intéressant de s'interroger sur les conditions de production des discours porteurs de ces unités en considérant qu'il pourrait y avoir un lien entre les situations dans lesquelles émergent les interactions et leur présence/absence.

[2] D'emblée, si, à l'instar de l'item le plus 'standard' de la série, *et caetera*, il est question d'un renvoi à des éléments implicitement suggérés, alors on peut convenir que les situations de communication dans lesquelles ils sont particulièrement nombreux sont le fruit d'une combinaison de paramètres qui permet aux interactants un partage suffisant d'informations pour l'interprétation des énoncés. De fait, la considération des unités traitées ici devrait intégrer la caractérisation des situations de communication. Cette prise en compte du contexte de production (l'usage) dans l'analyse des données me conduit à interroger la façon dont les corpus donnent accès aux métadonnées : dispose-t-on de métadonnées suffisantes pour tenter de mettre en lumière le degré de connivence des interactants ?

[3] Je vise ainsi une analyse d'éléments de la langue, en tenant compte de ce qui relève de la mise en parole. Dans un premier temps, il s'agit de montrer que cette entreprise permet d'éviter la multiplication des points de vue en se situant à un niveau qui les prendrait tous en charge. C'est ainsi que j'aboutis à la proposition du terme général *restituteurs d'ensemble par inférences* (REPI), en lieu et place d'autres, chacun motivé par un aspect de la fonction des unités :

As yet there is no generally accepted term for referring to expressions of this kind. They have been variously referred to in the literature as set marking tags (Dines 1980), vague category identifiers (Channell 1994), approximators (Erman 2001), general extenders (Overstreet 1999), discourse extenders (Norrby and Winter 2002), extension particles (Dubois 1992) and more. (Cheshire 2007 : 156).

[4] Dans un second temps, après avoir présenté les fondements de l'approche théorique sur laquelle je m'appuie, je mets à l'épreuve les corpus *OFROM* (Avanzi, Béguelin & Diémoz 2016a, 2016b) et *MPF* (Gadet 2017a) afin de voir dans quelle mesure les choix méthodologiques (donc théoriques) à l'origine de leur constitution et le mode d'accès aux données, sont favorables à une telle perspective. Il s'agit de faire dialoguer théorie et méthodologie, partant du principe que l'une et l'autre s'alimentent mutuellement (Gadet & Guerin 2012).

## 2 De l'intérêt du terme REPI

[5] Depuis que l'on s'intéresse aux données orales, et surtout que l'on en a

les moyens, on est en mesure d'étudier certains faits de langue absents, ou moins fréquents, à l'écrit. Ce n'est pas parce qu'on aurait affaire à deux langues, deux systèmes différents, comme le suggèrent les représentations communes, influencées par l'idéologie du standard qui repose sur une conception dichotomique du couple oral/écrit. Il en va, en fait, des conditions d'interaction des locuteurs lors d'un échange qui contraignent l'actualisation de la langue et sont donc vecteurs de variabilité. Ainsi, dans un certain nombre de situations de communication orales, les locuteurs ont la possibilité de partager (et de savoir qu'ils partagent) des informations qu'il n'est dès lors plus utile de mentionner explicitement, là où, souvent à l'écrit, ce n'est pas toujours garanti du fait notamment de la séparation des cadres spatial et temporel de la production et du mode de réception. Les unités sur lesquelles portent cette étude sont effectivement des éléments de la langue que l'on retrouve principalement et fréquemment à l'oral, sans exclure qu'elles peuvent apparaître dans des productions écrites, si les paramètres situationnels permettent d'atténuer la distance physique et symbolique entre les interactants. On pense par exemple aux messageries instantanées qui, en évacuant la contrainte de la transmission différée, permettent une collaboration des interactants telle qu'elle peut être observée dans de nombreuses situations d'oral. Un cas particulier est celui de la représentation d'un dialogue oral dans un écrit qui peut, là aussi entraîner la présence à l'écrit desdites unités. Cependant, leur présence n'est pas due au fait d'une forte connivence des interactants réels (auteur et lecteur) puisque, pour être interprétables, l'auteur doit fournir explicitement, hors du dialogue, les éléments qui permettront au lecteur de se représenter l'ensemble des savoirs implicites sur lequel repose l'interaction fictive.

[6] L'ensemble d'unités, a priori hétérogène, que je range sous l'étiquette REPI s'illustre dans la série d'énoncés suivante<sup>1</sup>, extraite des corpus *OFROM* et *MPF* constituée avec le souci d'illustrer des situations de communication et des profils de locuteurs différents<sup>2</sup> :

- (1) Mais si je suis en train d'écrire un mail *ou comme ça* j'ai pas forcément envie qu'elle vienne lire. (*OFROM*, unifr11-esb)

1 Pour faciliter la lecture, les transcriptions sont reproduites avec un aménagement des conventions : seuls les pauses sont marquées avec (.) et les segments inintelligibles avec X. Par ailleurs, je fais le choix ici de ponctuer les énoncés simplement en les faisant débiter par une majuscule et finir par un point ou point d'interrogation, le cas échéant. Les énoncés originaux, segmentés selon les conventions propres aux deux projets, peuvent être présentés ici sous une forme tronquée pour alléger le texte, avec le souci de ne pas nuire à l'interprétation des unités sur lesquelles porte l'analyse. En revanche, l'orthographe est fidèle à celle de la transcription d'origine. Ainsi, il peut y avoir des différences entre les énoncés issus d'*OFROM* ou de *MPF*, notamment parce que le premier admet les 'trucages orthographiques' et non le second.

2 Les sites internet des deux corpus permettent, à partir des références données à la suite de chaque énoncé, d'accéder à des informations concernant les locuteurs et les situations. Dans la suite du texte, comme annoncé, ces informations seront discutées.

- (2) Peut-être que j'ai trop tendance à me plaindre *ou comme ça* je sais pas. (OFROM, unine15-039)
- (3) Sur un site comme le Machu Picchu *ou comme ça* y a moins beaucoup moins de monde le matin. (OFROM, unine13-haa)
- (4) Je pense pas dans le centre de la Thaïlande bon aussi il y a beaucoup de prostitués *et tout ça*. (OFROM, unine15-931)
- (5) Pis elle m'a un peu motivée aussi à lire euh des livres *et tout ça*. (OFROM, unine16-021)
- (6) Bon après c'est parce que c'est pour l'unicef *et tout ça*. (OFROM, unine17-002)
- (7) C'est hyper romantique *et caetera* alors bon ben j'ai ri pis j'ai emmené. (OFROM, unine15-013)
- (8) Parce que leur but c'est de collectionner de peindre *et caetera*. (OFROM, unine15-013)
- (9) C'est presque une fois par mois mec t'as des gars tu sais pas sortis d'où qui arrivent qui t'apportent un petit croissant un verre de jus d'orange à ton à ta place de travail mec (.) à ton bureau tu sais t'es là eh bonjour ah c'est pour le nouvel abonnement qu'on a on fait *nin nin nin X* une petite action alors tous les collaborateurs ils ont droit à un jus d'orange pis un croissant. (OFROM, unifr11-maa)
- (10) Ils auraient coupé les vingt dernières minutes ce ce film serait été extraordinaire (.) pourquoi ils ont besoin à la fin de dire (.) voilà alors elle c'est la fille de parce que *nin nin nin* et pis il lui est arrivé ça et pis. (OFROM, unine11-jva)
- (11) Ils étaient tellement méchants (.) je me suis dit est-ce que c'est à cause de ça qu'ils sont devenus méchants (.) parce que moi j'ai pas envie de devenir comme eux (.) donc euh pff (.) au final j'ai lâché l'affaire (.) et parce que en plus ça coûte cher d'être dentiste parce que il faut acheter le matériel *et tout nanana*. (MPF, Nacer1)
- (12) Non mais des fois à la mini entreprise euh i- (.) des fois il dit des choses intéressantes *et tout* mais des fois i- (.) part en cacahuète. (MPF, Wajih1)
- (13) Pendant que j'étais au taf (.) au téléphone tout c'était impossible de d- de parler comme ça c'était impossible (.) tu avais que des vieux *et tout* (.) vas-y. (MPF, Marion2)
- (14) Ta ville c'était ut- c'était une petite ville donc c'est peut-être pour ça que tu pouvais traîner sans adultes *et tout ça*. (MPF, Roberto3a)
- (15) Une formation de six mois oui oui oui il y a une formation (.) nan c'est p-six mois oui si si six mois c'est (.) six mois c- non mais c'est vrai parce qu'en fait déjà avant l'avoir c'est une fois quand on a l'agrément parce que déjà il y aura la puéricultrice qui va passer pour voir le les lieux *et caetera*. (MPF, Sahar3)
- (16) Elle m'avait appelée avant elle me disait oui il y en a ils parlent sur toi *et*

- tout* mais je te le dis pas euh je te dis pas qui parce que voilà quoi. (MPF, Elodie2)
- (17) Et euh du coup alors à l'époque où elle était que que avec euh où elle était du côté obscur (rire) j'imagine qu'elle parlait créole *et tout* du coup enfin peut-être. (MPF, Emma3a)
- (18) Mais ça dépend des personnes aussi si direct vous venez ouais tu es un délinquant *nanana* comme la police aussi c'est normal qu'ils se révoltent. (MPF, Anna2)
- (19) Est-ce que est-ce que tu as remarqué ça mon ma génération toi tu es arrivé après mais dans notre génération on a toujours eu nos mots *et caetera* mais on était plus quand même très orientés verlan euh les euh les mots du romani parce que on est à côté de Montreuil *et caetera* il y avait pas autant de variété que maintenant. (MPF, Sandrine1)
- (20) Alors que d'autres (.) ben celles qui sont blanches de peau justement essaient de se foncer la peau (.) pour être euh bronzées *et caetera et caetera* (.) il y a quelque chose parfois qui est un peu perturbant. (MPF, JD3)

[7] À l'image de Cheshire (2007), citée précédemment, Béguelin & Corminboeuf (2017 : ¶ 1), s'intéressant en particulier à *ou comme ça* et *machin*, parviennent au même constat de l'hétérogénéité terminologique : « ils fonctionnent comme « extenseurs », « particules d'extension » (Ferré, 2009 ; angl. « general extenders », Overstreet, 2005), ou encore « clôtureurs » (Kahane et Pietrandrea, 2012), « prolongateurs » (Gadet, 2017) ». En s'intéressant aux travaux cités par les auteurs, on parvient à mettre en relation les termes utilisés et les approches et intérêts des travaux dans lesquels ils apparaissent. Autrement dit, si différents termes semblent recouvrir un même fait c'est qu'il est chaque fois question d'envisager le dit fait avec des approches, des niveaux d'analyse et des perspectives différentes. Lorsque Overstreet & Yule (1997 : 251) abordent la question en termes de *general extenders*, c'est dans la perspective d'une description qui satisfait des préoccupations d'ordre syntaxique : « General extenders have nonspecific reference or 'general' reference and (...) extend otherwise grammatically complete utterance ». On explique ainsi la capacité de ces unités à combler syntaxiquement un énoncé. Ce qui est suggéré avec *prolongateurs* va dans le même sens, à la différence que le terme ne laisse pas entendre l'absence de spécification référentielle. Cette remarque n'est pas anodine puisque s'en tenir à une référence *générale, non spécifique*, ne permet pas de s'interroger sur les raisons qui poussent les locuteurs à utiliser tel ou tel *general extender*. Or, quelle qu'en soit la conclusion, le fait que dans un énoncé tel que (11), on puisse retrouver les occurrences de deux d'entre eux à la suite devrait conduire, a minima, à se demander quel en est l'effet. De la même façon, parler de *clôtureurs* s'entend dans la perspective d'une description à visée systématisante, pour ne pas dire *automatisante*. Il s'agit d'un repérage permettant le balisage des discours : la présence d'un clôtureur permettrait, par

exemple, de signaler le terme d'une intention communicative.

Le dernier type d'éléments que l'on trouve dans les entassements ont la propriété de pouvoir clore un entassement, comme *et caetera* (...). Nous les appellerons des clô-  
 tureurs. Ils ne peuvent occuper que la dernière couche d'un entassement. (Kahane &  
 Pietrandrea 2012 : 1812).

[8] Dans tous les cas, les effets de l'oral sont intégrés à l'analyse par la re-  
 prise du modèle descriptif, initié par Blanche-Benveniste et al. (1990), qui permet  
 de concevoir le croisement des axes syntagmatique et paradigmatique. Partant, les  
*general extenders*, *continueurs* ou *clôtureurs* suggèrent une liste se déroulant  
 verticalement, à un endroit précis de l'axe horizontal, sur lequel se déploie l'énon-  
 cé, de façon linéaire. Cette schématisation facilite la lecture et, dans une certaine  
 mesure, la compréhension de l'organisation syntaxique. Dans (3), *ou comme ça*  
 permet de se représenter une liste, amorcée par *le Machu Picchu*, constituée  
 d'items référant à différents lieux touristiques, sous réserve d'une collaboration  
 suffisante des interactants, c'est-à-dire que la connaissance d'autres sites péruviens  
 équivalents soit partagée.

- (3a) Sur un site comme le Machu Picchu y a moins beaucoup moins de monde le  
 matin.
- (3b) Sur un site comme la ville sacrée de Chan Cha y a moins beaucoup moins  
 de monde le matin.
- (3c) Sur un site comme le Canyon du Colca y a moins beaucoup moins de  
 monde le matin.
- (3d) Sur un site comme ...

Ici, on suppose des référents commutables. En est-il de même dans les énoncés (2)  
 et (17)? La restitution d'une liste d'items commutables amorcée par *ou comme ça*  
 en (2) et *et tout* en (17) semble moins évidente. En (2), il n'est pas exactement  
 question d'activer une série d'actions appartenant au champ lexical de la plainte  
 (par exemple, *pleurer, gémir, sangloter...*). *Ou comme ça* semble suggérer l'acti-  
 vation de l'ensemble des éléments qui accompagne la plainte telle que vécue par le  
 locuteur. On est dès lors dans un processus plus cumulatif que substitutif. De la  
 même façon, en (17), il ne s'agit pas de la possibilité de substituer le créole à  
 d'autres langues. *Et tout* suggère d'activer tous les autres traits qui caractérisent  
 l'expression d'une identité antillaise, selon le locuteur (par exemple, écouter un  
 certain style de musique, porter un certain style de vêtements, ...).

[9] En (3) comme en (2) et (17), il est question d'une procédure inférentielle  
 rendue possible par le partage de certaines connaissances en lien avec l'élément  
 déclencheur explicitement mentionné. Dans tous les cas, on vise la restitution d'un  
 ensemble d'items qui, mis en commun, dessinent un tout qui constitue finalement

le référent évoqué. On s'approche ainsi de la notion d'*approximators* proposée par Erman (2001 : 1341) : « Through approximators the speaker gives the listener/s "a rough but sufficiently exact idea about a certain state of affairs for the general purpose of the conversation" ». En (3), la liste d'items commutables permet au locuteur de ne pas activer le Machu Picchu (ou un autre des items de la liste) pour ses propriétés historiques, archéologiques ou géographiques mais pour ce qu'il a en commun avec tout autre site touristique. En (2), de la même façon, c'est la mise en commun de tous les faits, gestes et attitudes du locuteur lorsque celui-ci se plaint qui permettent de suggérer un état en particulier et non un fait, geste ou attitude. Enfin, en (17), c'est l'identité antillaise, émergeant de la mise en réseau des items récupérables à partir de l'élément déclencheur explicite (*le créole*), qui est visée et non un trait caractéristique en particulier. En envisageant ainsi le sens procédural de *ou comme ça, et tout, ...*, la représentation sous forme de liste verticale d'items commutables, vraisemblablement opérante pour les besoins d'une description strictement syntaxique, est moins adaptée pour aborder, comme je le propose, la spécificité sémantique d'unités telles que celles traitées ici. En effet, les éléments du paradigme ne sont pas interchangeable mais doivent être pensés comme un ensemble à partir duquel émerge un référent conceptuel.

[10] Sans chercher à entretenir le foisonnement terminologique, je propose ainsi de parler de *restituteurs d'ensemble par inférence* (REPI). Il s'agit d'actualiser la proposition, à mon sens trop peu exploitée, de Dines (1980 : 22) qui parle de *set marking tags*, définis en ces termes :

In every case their function is to cue the listener to interpret the preceding element as an illustrative example of some more general case. Tags, then, operate on "parts" to relate them to "wholes". (...) The presence of a clause-terminal tag indicates that an underlying general notion has been realised by a specific example.

S'il m'apparaît nécessaire de recourir à ce terme c'est que les travaux s'appuyant sur les catégories traditionnelles, établies à partir de productions qui ne sont pas issues de situations de communication permettant une forte collaboration communicationnelle, aboutissent à la reconnaissance de l'inadaptation de ces catégories. C'est, par exemple, ce qu'on comprend en substance des propos de Bilger (1989 : 106) lorsqu'elle conclut son étude sur *et tout ça* :

Cette forme élimine donc les frontières des catégories ; elle permet de mettre sur le même plan forme nominale et forme verbale (la partie notionnelle du verbe). Elle n'est pas qu'un indice discursif proche des phatiques puisque c'est apparemment la seule forme qui nous permette de laisser ouverte et non déterminée cette place syntaxique occupée par le verbe constructeur.

[11] Il s'agit donc, à la lumière des données accessibles aujourd'hui, d'envisager de nouvelles catégories, un autre niveau d'analyse où les parties du discours

ne sont pas déterminantes quant à la catégorisation. On préfère l'idée d'ensemble plutôt que liste afin de ne pas limiter à des équivalents de l'élément déclencheur les éléments implicites à restituer. Ces derniers doivent être constitutifs d'un tout accessible, de la même façon par les interactants. En (1), *ou comme ça* appelle la restitution de toute autre situation caractérisée par ce que *je* suppose partager avec *tu* à propos *d'écrire un mail*. En (4), *tout ça* appelle la restitution de tout ce que *je* suppose partager avec *tu* à propos de l'aspect *décadent* de la Thaïlande. En (13), *et tout* appelle la restitution de ce que *je* suppose partager avec *tu* à propos des *vieux*, leurs attitudes et leurs façons de penser. En (7), *et caetera* appelle la restitution de ce que *je* suppose partager avec *tu* à propos de ce à quoi un contexte romantique peut conduire. En (19), *et caetera* appelle la restitution de ce que *je* suppose partager avec *tu* à propos de la culture *jeune* qui s'illustre notamment à travers des pratiques langagières. En (10), *nin nin nin* appelle la restitution de ce que *je* suppose partager avec *tu* à propos de l'histoire racontée par le film. Enfin, en (11), *nana-nan* appelle la restitution de ce que *je* suppose partager avec *tu* à propos des dépenses impliquées dans la poursuite d'études de dentiste.

### 3 Quelle approche de la variation ?

[12] L'étude des REPI peut s'ancrer dans une réflexion épistémologique sur la variation linguistique. En effet, les tenants de l'approche la plus courante de la variation, l'approche variationniste, ne permettent pas de traiter efficacement de telles unités : les études quantitatives ne révèlent pas de corrélation évidente entre la présence de REPI et les critères socio-démographiques traditionnellement convoqués. Il n'y aurait pas de profils sociaux ni de situations (caractérisant celles-ci relativement à un dit degré de formalisme) favorisant particulièrement la présence de REPI. Comme le souligne Cheshire (2007 : 188) :

Other than the difference in some forms of the adjunctives, there was no social class or gender variation in the uses of the general extenders. However this does not preclude their being involved in a more complex type of sociolinguistic variation, if other features with similar pragmatic functions were included in the analysis. I have stressed that discourse variation differs from phonological or morphosyntactic variation in that speakers are not obliged to use a given discourse form.

Cette affirmation admet, en substance, les limites de l'approche variationniste qui n'intègre pas les effets de l'interaction.

[13] Or, sans une prise en compte de ce qui se joue entre les interactants au moment de l'échange, la compréhension du fonctionnement des REPI, de la pertinence de leur présence/absence, ne peut être que partielle, voire impossible. Ce constat n'est pas récent puisqu'il était déjà formulé par Dines (1980 : 29) :

The outcome of the preceding exercise is that it is as yet pointless to undertake quan-



titative analysis, for the taxonomy of variants is demonstrably incomplete. The next stage of the inquiry involves examining style-shifting within individuals, self-corrections and reformulations and an interactive analysis of continuing discourse as well as considering some psychological and pragmatic factors. I find myself agreement then with Lavandera, who does not argue in principle against the extension of the model – but points out that in our present state of knowledge quantitative research may be unrevealing.

L'accès au *next stage* évoqué par Dines (1980 : 29) peut, par exemple, se concevoir dans le cadre théorique d'une approche communicationnelle de la variation (Guerin 2017) où l'on tente de décrire la variation comme l'effet de l'imbrication des facteurs d'ordre macro, au niveau social, et des facteurs d'ordre micro, au niveau de l'interaction. Autrement dit, appliqué à l'étude des REPI, ce cadre permet d'intégrer à l'analyse l'identité des locuteurs/scripteurs en interaction dans une situation donnée tout en considérant les normes et représentations communes. Cette perspective implique une représentation horizontale et décloisonnée de la variation qui ne soutient pas l'absolue pertinence d'une unique forme (le 'bon usage'). Si l'on ne nie pas le caractère socialement prestigieux de la forme standard et, par extension, l'influence de sa reconnaissance sur les pratiques (ce qui explique, par exemple, les hypercorrections), l'observation des pratiques réelles montre qu'en fait, à chaque situation correspond une norme qui s'établit compte tenu de la combinaison des paramètres qui la caractérise.

[14] On cherche donc à se demander quelles sont ces combinaisons favorisant l'usage des REPI. En l'occurrence, étant donné le sens procédural révélé de l'observation de données telles que celles présentées ici, on peut d'ores et déjà dire que l'usage des REPI requiert et/ou entretient une connivence suffisante des interactants pour pouvoir s'appuyer sur des implicites. Le partage de connaissances ne doit pas se limiter à un savoir communément partagé puisque la procédure interprétative conduit à une opération collaborative en deux temps :

- la restitution d'items déclenchée par un élément explicite, selon un lien tel que considéré par le locuteur ;
- l'activation d'un référent, fruit de la mise en réseau des items.

En somme, l'efficacité expressive des REPI dépend de la coopération des interactants. On voit ainsi qu'ils font partie de ces éléments de la langue qui ne peuvent être saisis sans une prise en compte des conditions de production.

[15] Il n'est pas surprenant de remarquer que *ça* entre dans la composition de certains d'entre eux (*et tout ça*<sup>3</sup>, *ou comme ça*). En effet, Cadiot, (1988), Achard (2001) ou encore Guerin (à paraître) ont montré que le sens procédural de *ça* implique une certaine collaboration des interactants, possible à la condition d'une connivence suffisante : « *Ça's* presence signals the cognitive manipulation (summarization, abstraction, inferencing) of information already available in the

---

3 Voir Johnsen (2011, 2019) ou encore Bilger (1989).

context or in the hearer's world knowledge » (Achard 2001 : 56). *Ça* seul ou constituant d'un REPI, contrairement aux pronoms (bien que la grammaire traditionnelle le range dans cette catégorie, notamment parce qu'elle l'envisage comme étant l'équivalent de *cela*), n'a pas pour sens d'activer un référent identifiable en cotexte ou en contexte. Il active ce que l'on peut penser, inférer, d'un référent donné, ce qui suppose nécessairement que les interactants partagent non seulement la connaissance de ce référent mais aussi ce que l'on peut en penser, en inférer.

*Ça* ne peut être considéré comme une forme de reprise (anaphore) à l'identique, de cet antécédent supposé, mais bien plus comme la trace d'un accès propositionnel à sa référence (à ce dont l'antécédent « nous parle » dans un énoncé donné). Même lorsqu'il y a bien reprise, l'essentiel est que cette reprise (considérée ici comme un processus) ne traite pas son antécédent comme une « valeur », mais comme l'argument d'une ou plusieurs proposition(s) implicite(s), plus ou moins accessibles et variablement explicitables. (Cadiot 1988 : 174)

[16] À l'image du fonctionnement de *ça*, les REPI ne renvoient pas strictement à l'ensemble des items restituables, constituant l'ensemble suggéré par l'élément explicite déclencheur. Cet ensemble est l'argument d'une ou plusieurs proposition(s) implicite(s), celle(s) qui condui(sen)t à identifier un référent tel qu'envisagé par le locuteur. Dans (5), *et tout ça* appelle dans un premier temps la restitution d'items déclenchée par *lire des livres*. En l'occurrence, on peut imaginer qu'il s'agit des actions, attitudes et comportements envisagés par le locuteur comme caractéristiques d'une certaine conformité aux attendus du prestige social et notamment scolaire. Cette série d'items constituée en ensemble entraîne, dans un second temps, l'activation d'un référent unique, une attitude générale et non des actions, attitudes et comportements envisagés de façon dissociés.

#### 4 Les REPI dans *OFROM* et *MPF*

[17] Cette considération du sens procédural des REPI peut (devrait?) avoir des retombées sur la façon de constituer et de traiter des corpus de données. En effet, étant donné qu'il est peu probable qu'ils apparaissent lorsque les interactants n'entretiennent pas une connivence suffisante, on cherche alors à retenir des données issues d'enregistrements dans des situations dont la combinaison des paramètres le permet. Cela implique d'être en mesure d'évaluer les situations d'interaction enregistrées. Cette entreprise conduit à tenter de mettre en lumière les paramètres situationnels, ceux-ci relevant autant du niveau macro (social) que du niveau micro (interaction) (Baude & Guerin 2019). Pour tenter d'accéder à ces paramètres on a recours aux métadonnées, censées renseigner sur les conditions et le contexte de l'interaction et sur l'identité des interactants.

[18] Tous les corpus rendus disponibles proposent une série de métadonnées associées aux données. Cependant, tous n'en proposent pas le même degré de gra-

nularité. Le corpus *OFROM* donne accès à des informations sur l'identité des locuteurs :

Les enquêteurs avaient pour consigne de recueillir, pour chaque locuteur enregistré, un certain nombre d'informations qui devaient permettre de trier les locuteurs selon (...) qu'ils sont francophones natifs (L1) ou non (L2), selon leur niveau socio-éducatif, selon qu'il s'agit d'un homme ou d'une femme. On peut également indiquer un intervalle temporel pour spécifier la tranche dans laquelle le locuteur est né. (Avanzi, Béguelin & Diémoz 2016b : 10-11)

Ces informations sont rudimentaires, pour reprendre le terme des auteurs du document de référence du projet. En fait, elles permettent de situer socialement les individus, selon une représentation catégorielle, dans la tradition des études variationnistes. Concernant l'identification des paramètres relatifs aux types de situations d'interaction, très peu d'informations sont fournies. On sait qu'il s'agit d'entretiens menés par des étudiants, que, dans la plupart des cas, les locuteurs enregistrés se livrent à un monologue et « une plus petite partie des enregistrements ressemblent davantage à des interactions, puisqu'ils impliquent au moins deux personnes qui parlent à bâtons rompus » (Avanzi, Béguelin & Diémoz 2016b : 4)<sup>4</sup>. Partant, il apparaît difficile de s'appuyer sur ces informations pour déterminer le degré de connivence des interactants (enquêteur et informateur), notamment parce que rien n'est dit quant à la relation de ces derniers, leur histoire conversationnelle, les savoirs, connaissances et expériences partagés.

[19] Au contraire, dans le corpus *MPF*, c'est

la relation entre enquêteur et enquêté qui a été mise au principe du recueil. C'est pourquoi le critère essentiel retenu pour constituer le corpus a été la *proximité*<sup>5</sup> des protagonistes (voir Gadet et Guerin 2012), et l'existence entre eux d'une *histoire conversationnelle*. (Gadet 2017b : 16)

De fait, les métadonnées fournies par *MPF* renseignent au-delà des caractéristiques socio-démographiques et tentent de rendre compte du niveau de *proximité*, autrement dit de connivence, des interactants. Naturellement, ces informations peuvent être de natures différentes et difficilement systématisables. Contrairement à ce que permet *OFROM*, *MPF* ne peut pas proposer de concordancier qui permettrait de générer des données relativement à ce niveau de granularité des métadonnées. On est, de fait, dans une approche nécessairement qualitative.

[20] En somme, si les deux corpus offrent un accès à des données illustrant un aspect de la langue française sans prétention de représentativité, cette perspec-

4 On peut être surpris de voir le mot *interaction* réservé à ces discussions, comme s'il n'y avait pas d'interaction lorsque l'informateur produit un monologue.

5 La notion de *proximité* évoquée par Gadet renvoie au continuum *proximité - distance* proposé par Koch & Oesterreicher (2001).

tive étant en soit questionnable (Gadet & Wachs 2015), ils diffèrent, d'une part, sur le plan des situations de communication enregistrées : dans *OFROM*, les interactions tendraient, a priori, davantage vers la distance communicationnelle (Koch & Oesterreicher 2001), bien qu'il y ait probablement divers degrés de distance entre les interactants, du fait du format entretien, puisque familiarité des enquêteurs et des informateurs n'est pas une condition méthodologique ; dans *MPF*, la majorité des enregistrements, tendent, a priori, davantage vers la *proximité communicationnelle* puisqu'il s'agit soit d'enregistrements dits *écologiques*, entre pairs sans intervention de l'enquêteur, soit d'entretiens dits *de proximité*, c'est-à-dire où enquêteurs et informateurs ont une familiarité préalable au moment de l'enregistrement. D'autre part, les métadonnées accessibles ne sont pas du même ordre pour les deux corpus étant donné que, notamment, aucune information n'est fournie concernant la relation enquêteurs-informateurs dans *OFROM*.

[21] Ces constats faits, quelle conséquence pour l'étude des REPI ? D'emblée, compte tenu de ce qui a été vu de leur sens procédural et des opérations nécessaires à leur interprétation, on peut être tenté de supposer qu'il y aurait davantage d'occurrences dans *MPF*. Pour tester cette hypothèse, j'ai relevé toutes les occurrences de *et tout (ça), ou comme ça, nanana / ninninnin, et caetera*, dans les deux corpus<sup>6</sup>. En effet, il y a nettement plus d'occurrences de REPI dans *MPF*, soit une moyenne de 0,2 occurrence pour cent mots par enregistrement alors que dans *OFROM*, la moyenne s'élève à 0,06. On est donc tenté de confirmer l'hypothèse selon laquelle l'emploi des REPI est relatif à la proximité communicationnelle qui favorise la connivence des interactants. Néanmoins, si l'on corrèle le sens procédural des REPI à ce qui se joue entre les interactants au moment de l'échange, ce niveau d'approche des données n'est pas encore suffisamment précis, notamment parce qu'il masque le fait qu'il y a des enregistrements contenant peu ou pas de REPI dans *MPF* et, inversement, qu'il y a des enregistrements qui en contiennent un nombre significatif dans *OFROM*<sup>7</sup>. C'est en s'intéressant aux métadonnées qu'on est en mesure de l'expliquer.

[22] Dans *MPF*, l'enregistrement contenant la plus forte concentration de REPI (1,1% de l'ensemble des mots) concerne une situation d'interaction caractérisée par les paramètres suivants :

- Une conversation libre (par opposition à un entretien) ;
- L'échange se tient au domicile de l'informateur ;
- Les interactants sont de 'bons amis' et voisins ;
- La thématique abordée est en lien avec des expériences ou personnes connues des deux interactants, comme l'illustre la présence des *tu vois* dans

6 Au moment du relevé, *OFROM* comptait un million de mots et *MPF* 1052974 mots. Cette proximité numérique n'assure qu'une certaine comparabilité au niveau de la densité des corpus.

7 On n'ignore pas le fait que la présence/absence de REPI peut être liée aux habitudes langagières des locuteurs, certains locuteurs privilégiant certains REPI ou utilisant d'autres unités verbales ou non verbales pour assurer leur fonction.

l'extrait suivant qui, en plus de sa fonction phatique, suggère que les informations sont bien connues des deux interactants :

- (21) C'est que apr- parce qu'après elle donc du coup le mec du coup le mec euh moi je la chope et tout tu vois donc on recommence notre petit cinéma nos petits euh affaires habituelles tu vois. (*MPF*, Adeline2)

La combinaison de ces paramètres favorise à l'évidence la connivence des interactants. Conséquemment, les productions s'élaborent en s'appuyant sur une somme conséquente d'informations implicites. Il n'est pas utile ici d'explicitier les *petites affaires habituelles*.

[23] Si l'on s'intéresse à l'enregistrement contenant le moins de REPI dans *MPF* (aucun), on peut caractériser la situation d'interaction à partir des paramètres suivants :

- Une conversation libre (par opposition à un entretien) ;
- L'échange se tient au domicile de l'informateur ;
- Les interactants sont frères et en présence de la compagne de l'un des deux ;
- La thématique abordée est celle des jeux vidéo. Plus précisément, il est question de la description technique du graphisme d'un jeu.

On voit que ce qui distingue la première situation de la seconde se situe au niveau de la thématique abordée. Outre le genre (Biber & Conrad 2009) du discours qui n'est pas le même, dans la seconde interaction, il est question d'introduire un élément informatif nouveau, les caractéristiques techniques du graphisme du jeu. Bien que les autres paramètres de la situation favorisent la connivence, cet élément pose nécessairement une certaine distance (voir Koch & Oesterreicher, 2001). En effet, il n'est dès lors plus possible d'élaborer les productions en supposant un partage de connaissances. Autrement dit, on ne peut s'appuyer sur des implicites sous peine de ne pas être compris. Si *MPF* n'offrait pas une possibilité d'accès à des métadonnées d'un tel niveau de granularité, on ne serait pas en mesure d'expliquer la forte présence et l'absence de REPI dans deux enregistrements d'interactions dont les caractéristiques plus générales semblent similaires.

[24] Dans *OFROM*, l'accès aux métadonnées et leur nature permettent de caractériser la situation d'interaction concernant l'enregistrement contenant la plus forte concentration de REPI (0,37% de l'ensemble des mots) selon les éléments suivants<sup>8</sup> :

---

<sup>8</sup> Les tableaux sont ceux accessibles sur le site du projet.

Information sur le locuteur	
Nom	unine13-bea
Langue	Français L1
Sexe	Femme
Année de naissance	1985
Lieu de naissance	Genève
Lieu de résidence	Genève
Canton de résidence actuel	Genève
Y habite depuis	1985
Niveau socio-éducatif	Niveau 4
Occupation	Étudiante
Âge au moment de l'enregistrement	27
Statut familial	Célibataire

Informations sur l'enregistrement	
Transcription	unine13a03d
Qualité	Bonne
Type	d
Genre de parole	discussion
Uni	UNINE
Propriétaire	Mathieu Avanzi
Lieu d'enregistrement	Genève
Canton	Genève
Enregistré par	Pauline Dubosson
Transcrit par	Maud Erdinger
Révisé par	François Delafontaine
Date de l'enregistrement	16/04/2012
Date de révision	03/06/2014
Date de mise en ligne	22/06/2018

Tableau 1: Métadonnées de l'enregistrement contenant la plus forte concentration de REPI

[25] Les tableaux suivants concernent la situation d'interaction concernée par l'enregistrement contenant le plus faible taux de REPI (0,01% de l'ensemble des mots) :

Information sur le locuteur		Informations sur l'enregistrement	
Nom	unifr11-esb	Transcription	unifr11a01d
Langue	Français L1	Qualité	Bonne
Sexe	Femme	Type	d
Année de naissance	1996	Genre de parole	discussion
Lieu de naissance	Essert	Uni	UNIFR
Lieu de résidence actuelle	Essert	Propriétaire	Alain Berrendonner
Canton de résidence actuel	Fribourg	Lieu d'enregistrement	Essert
Y habite depuis	1996	Canton	Fribourg
Niveau socio-éducatif	NR	Enregistré par	Maëlle Andrey
Occupation	Élève de 3ème du CO	Transcrit par	Maëlle Andrey
Âge au moment de l'enregistrement	15	Révisé par	François Delafontaine
Statut familial	Célibataire	Date de l'enregistrement	24/01/2011
		Date de révision	01/12/2012
		Date de mise en ligne	21/06/2018

Tableau 2: Métadonnées de l'enregistrement contenant la plus faible concentration de REPI

Ces informations pourraient conduire à tirer des conclusions quant à une relation entre l'âge des locutrices ou leur lieu de naissance et la présence/absence des REPI, puisque seuls ces éléments diffèrent dans les deux situations. Or, si l'on regarde les enregistrements dans lesquels on relève la plus forte concentration de REPI après celui pris en exemple, on s'aperçoit que les profils sont variables. Ainsi, les métadonnées qui accompagnent les données du corpus *OFROM*, réduites à des critères sociolinguistiques minimaux, pour reprendre les termes du document de référence du projet, ne permettent pas une approche du fonctionnement des REPI suffisamment affinée pour aboutir à la mise en lumière de leur sens procédural tel qu'il a pu être proposé dans cette étude.

## 5 Conclusion

[26] Ce que nous permettent les nouvelles techniques de recueil et de traitement des données, notamment orales, fait apparaître des faits de langue qui, faute de pouvoir pleinement observer leur fonctionnement, ont longtemps été ignorés de la description linguistique. Dans la plupart des cas, on les a considérés comme relevant de variétés fautives ou si peu prestigieuses que leur absence de traitement ne souffrait d'aucune contestation. De fait, l'appareillage descriptif traditionnel ne propose pas une grille de lecture adaptée à une analyse intégrant la spécificité des

situations dans lesquelles lesdits faits de langues sont produits. Cet état de fait contribue à maintenir une approche de la variation qui ne prend pas en charge un certain nombre de facteurs et dans laquelle on n'est finalement pas en mesure de se détacher totalement d'une représentation qui suppose un 'bon usage', une norme et ses variantes.

[27] En se situant au-delà des seules catégorisations socio-démographiques traditionnellement convoquées et en intégrant des considérations relevant de l'interaction en situation, les cartes sont redistribuées et les faits de langue peuvent être observés et analysés pour ce qu'ils sont, sans référence à ce que l'on peut en dire lorsqu'ils apparaissent ou non dans un type de situation en particulier. Cette approche révisé par conséquent la ligne de partage oral/écrit puisque les phénomènes tels que les REPI n'ont pas lieu d'être envisagés exclusivement à l'oral, dès lors que la situation permet une connivence suffisante (Guerin & Moreno 2016).

[28] Il s'agit ainsi d'impliquer les métadonnées dans l'analyse des données. Celles-ci devraient ainsi être pensées comme indispensables à la compréhension du fonctionnement des unités de la langue. Cela invite à mener une réflexion sur leur nature et la façon dont il serait possible de les évaluer dans une perspective davantage quantitative (Baude & Guerin 2019).



## Bibliographie

- Achard, Michel 2001. French *ça* and the dynamics of reference. *LACUS Forum* 27, 49-62. <https://archive.org/details/lacus27-new/page/n47/mode/2up>.
- Avanzi, Mathieu, Marie-José Béguelin, Federica Diémoz 2016a. De l'archive de parole au corpus de référence : la base de données orales du français de Suisse romande (OFRON). *Corpus* 15. <http://journals.openedition.org/corpus/3060>.
- Avanzi, Mathieu, Marie-José Béguelin, Federica Diémoz 2016b. *OFRON. Corpus oral de français de Suisse Romande*. [http://www11.unine.ch/uploads/Documents/AM-MJB-FD\\_OFRON.pdf](http://www11.unine.ch/uploads/Documents/AM-MJB-FD_OFRON.pdf).
- Baude, Olivier, Emmanuelle Guerin 2019. Questions méthodologiques pour construire la « ville » à partir des ESLOs. Françoise Gadet (éd.). *Les métropoles francophones européennes en temps de globalisation*. Paris : Garnier, 41-55.
- Béguelin, Marie-José, Gilles Corminboeuf 2017. *Ou comme ça, machin* et autres marqueurs d'indétermination dans les listes. *Discours* 20. <https://journals.openedition.org/discours/9275>.
- Biber, Douglas, Susan Conrad 2009. *Register, genre, and style*. Cambridge : Cambridge University Press.
- Bilger, Mireille 1989. Les réalisations en *et tout ça* à l'oral. *Recherches sur le français parlé* 9, 97-109.
- Blanche-Benveniste, Claire et al. 1990. *Le français parlé. Études grammaticales*. Paris : CNRS Éditions.
- Cadiot, Pierre 1988. De quoi *ça* parle ? À propos de la référence de *ça*, pronom-sujet. *Le français moderne* 56, 174-192. <https://en.calameo.com/read/0009039479cc0ab1d5f2f?authid=q9AQMxnvIM7X>.
- Cheshire, Jenny 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11, 155-193.
- Dines, Elizabeth R. 1980. Variation in discourse – "and stuff like that". *Language in Society* 9, 13-31.
- Erman, Britt 2001. Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics* 32, 1337-1359.
- Gadet, Françoise (éd.) 2017a. *Les parlers jeunes dans l'Île de France multiculturelle*. Paris : Ophrys.
- Gadet, Françoise 2017b. Introduction. Françoise Gadet (éd.). *Les parlers jeunes dans l'Île de France multiculturelle*. Paris : Ophrys, 15-26.
- Gadet, Françoise, Emmanuelle Guerin 2012. Des données pour étudier la variation : petits gestes méthodologiques, gros effets. *Cahiers de linguistique* 38, 41-65.
- Gadet, Françoise, Sandrine Wachs 2015. Comparer des données de corpus : évidence, illusion ou construction ? *Langage et société* 154, 33-49.
- Guerin, Emmanuelle 2017. Éléments pour une approche communicationnelle de la variation. Henry Tyne et al. (éds.). *La variation en question(s). Hommages à Françoise Gadet*. Bruxelles : Lang, 57-76.
- Guerin, Emmanuelle, à paraître. Une description fondée sur l'oral (?) : penser *ça* sans cela. Paul Cappeau (éd.). *Une grammaire sous l'influence de l'oral*. Rennes : Presses universitaires de Rennes.
- Guerin, Emmanuelle, Anaïs Moreno 2016. Le discours rapporté dans les interactions orales et écrites : au-delà d'une opposition de surface. Daniel Jacob, Françoise Gadet, Anthony Lodge (éds.). *Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Section 9 : Rapports entre langue écrite et langue parlée*. Nancy, ATILF/SLR, 61-72. <http://www.atilf.fr/cilpr2013/actes/section-9/CILPR-2013->

- 9-Guerin-Moreno.pdf.
- Johnsen, Laure Anne 2011. Un éclairage sur le fonctionnement référentiel de *tout ça* en fin de liste. Gilles Corminboeuf, Marie-José Béguelin (éds.). *Du système linguistique aux actions langagières. Mélanges en l'honneur d'Alain Berrendonner*. Bruxelles : De Boeck/Duculot, 487-505.
- Johnsen, Laure Anne 2019. *La sous-détermination référentielle et les désignateurs vagues en français contemporain*. Bern : Lang.
- Kahane, Sylvain, Paola Pietrandrea 2012. La typologie des entassements en français. Franck Neveu et al. (éds.). *CMLF 2012 - 3e Congrès mondial de Linguistique française, 1809-1828*. <https://doi.org/10.1051/shsconf/20120100238>.
- Koch, Peter, Wulf Oesterreicher 2001. Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit. Günter Holtus, Michael Metzeltin, Christian Schmitt (éds.). *Lexikon der romanistischen Linguistik. Band / Volume 1,2. Methodologie (Sprache in der Gesellschaft / Sprache und Klassifikation / Datensammlung und -verarbeitung) / Methodologie (Langue et société / Langue et classification / Collection et traitement des données)*. Tübingen : Niemeyer, 584-627.
- MPF = Françoise Gadet (éd.) 2010-2019. *Multicultural Paris French*. <https://mpfvitrine.modyco-fr/vitrine>.
- OFROM = Mathieu Avanzi, Marie-José Béguelin, Federica Diémoz (éds.) 2012-2019. *Corpus oral de français de Suisse romande*. <http://www.unine.ch/ofrom>.
- Overstreet, Maryann, George Yule 1997. On being inexplicit and stuff in contemporary American English. *Journal of English Linguistics* 25, 250-258.