



HAL
open science

Etude syntaxique, discursive, lexicométrique et sociolinguistique du mot quoi dans le Corpus du français parlé parisien des années 2000

Serge Fleury, Florence Lefevre, Mat Pirès

► To cite this version:

Serge Fleury, Florence Lefevre, Mat Pirès. Etude syntaxique, discursive, lexicométrique et sociolinguistique du mot quoi dans le Corpus du français parlé parisien des années 2000. Yana Grinschpun et Judith Nyee-Doggen (éds). Regards croisés sur la langue française : usages, pratiques, histoire. Mélanges en l'honneur de Sonia Branca-Rosoff, p. 97-112, 2012. halshs-01143334v2

HAL Id: halshs-01143334

<https://shs.hal.science/halshs-01143334v2>

Submitted on 21 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etude syntaxique, discursive, lexicométrique et sociolinguistique du mot *quoi* dans le Corpus du français parlé parisien des années 2000

Serge Fleury, Université Sorbonne nouvelle, Paris 3, SYLED (Clesthia)
Florence Lefeuvre, Université Sorbonne nouvelle, Paris 3, SYLED (Clesthia)
Mat Pires, Université de Franche-Comté, LLC-ELLIADD EA4661

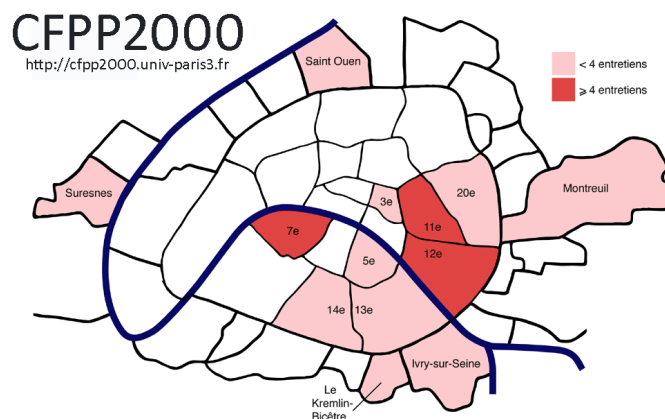
A citer sous cette forme : Fleury S., Lefeuvre F., Pirès M., 2012 : « Quoi dans le Corpus du français parlé parisien : étude syntaxique, lexicale et sémantique », Regards croisés sur la langue française : usages, pratiques, histoire. Mélanges en l'honneur de Sonia Branca-Rosoff (Y. Grinshpun & J. Nyee-Doggen eds), Paris, Presses de la Sorbonne Nouvelle, p. 97-112.

L'objet de cet article d'hommage à Sonia Branca est de proposer une étude syntaxique, discursive, lexicométrique et sociolinguistique du mot *quoi* dans le Corpus du français parlé parisien des années 2000 (CFPP2000). Ce corpus a été initié par Sonia Branca, et a bénéficié d'une Bourse de recherche de la Ville de Paris en 2006. Actuellement, il est soutenu par la DGLFLF. Un premier projet de recherche collectif « ANR » a été monté en janvier 2010, porté par Sonia Branca-Rosoff, et un deuxième projet ANR, porté par Florence Lefeuvre, a été soumis en octobre 2011. Nous présenterons, dans la première partie de cette étude, les outils disponibles sur le site du Corpus du français parlé parisien. Cela nous permettra de proposer, dans une deuxième partie, un cadre syntaxique d'étude du mot *quoi*, et dans une troisième partie, un regard sociolinguistique sur l'utilisation du mot *quoi*.

1. Données générales du Corpus du Français Parlé Parisien des années 2000

1.1. Le corpus CFPP2000

Le Corpus du Français Parlé Parisien des années 2000 (désormais noté CFPP2000) est une base de données sur le français parlé par des adultes de la région Ile-de-France. Le corpus, en cours de construction, et qui devrait atteindre un million de mots environ (une soixantaine d'heures), a été constitué à partir d'entretiens sur la relation des habitants à leur quartier. Il comprend actuellement environ 500 000 mots transcrits orthographiquement et alignés au tour de parole.



Les données du corpus CFPP2000 sont présentes en ligne sous différentes facettes. Les fichiers de transcriptions et les fichiers audios associés sont directement disponibles au téléchargement : les fichiers audio sont disponibles aux formats wav¹ et mp3, les fichiers de transcription sont au format Transcriber² ; on peut ainsi réutiliser conjointement ces données dans le logiciel Transcriber pour rétablir l'alignement entre le fichier de transcription et le fichier audio. Les fichiers de transcriptions et les fichiers audio alignés sont aussi directement accessibles en ligne de manière synchrone : la lecture de chaque alignement se faisant directement en ligne dans le navigateur de l'utilisateur, on utilise pour cela des ressources mises en œuvre par le Centre de ressources pour la description de l'oral³. Les données de transcription ont aussi été reformatées pour être disponibles en ligne dans un concordancier : les requêtes peuvent être faites sur les données brutes (le texte des transcriptions) ou sur une version étiquetée (via Cordial) des transcriptions. L'ensemble des fichiers de transcription est aussi disponible au téléchargement dans un format compatible avec les logiciels Lexico3⁴ et Le Trameur⁵ : on peut ainsi réaliser des opérations textométriques sur ces données.

Les données du corpus sont aussi associées à des métadonnées. Celles-ci permettent de définir un ensemble de descripteurs et les valeurs utilisées pour décrire et catégoriser le contenu et les objets des données visées. En règle générale, elles sont utilisées pour des opérations de recherche dans les données (sujet, auteur, date, mots clé etc.), pour diffuser les données et pour les archiver. On peut aussi concevoir des opérations complexes permettant de croiser des requêtes et d'interroger simultanément les métadonnées et les contenus : dans notre cas, on pourrait par exemple envisager de récupérer les données du corpus correspondant à une interrogation du type : « Quelles sont les fréquences d'utilisation du mot *quoi* interrogatif et de la périphrase en *qu'est-ce que* chez les personnes interrogées de moins de 30 ans ? ». Concrètement, les métadonnées permettent d'identifier les entretiens concernés par le critère d'âge, les identifiants des données résultantes devant permettre de calculer les fréquences visées.

1.2. Aspects quantitatifs CFPP2000

Le tableau ci-dessous présente l'ensemble des entretiens de CFPP2000 (durée de l'entretien, nombre d'occurrence de formes pour chaque entretien) :

¹ Le corpus est donc exploitable pour des spécialistes de phonétique et de prosodie.

² <http://trans.sourceforge.net/en/presentation.php>

³ <http://crdo.risc.cnrs.fr/exist/crdo/>

⁴ <http://www.tal.univ-paris3.fr/lexico/>

⁵ <http://www.tal.univ-paris3.fr/trameur/>

	IDENTIFIANT CFPP2000	Durée entretien (mn)	Nombre d'occurrences (segmentation Lexico3)
1	CFPP2000 03-01 OSOBI_HAK_H_32_118_3E	46	10 868
2	CFPP2000 05-01 CHRISTOPHE_ANDRÉ_H_02 JIMME_AIMÉ_ANDRÉ_F_03	74	16 617
3	CFPP2000 07-01 LUCRÉCE_LÉONORE_F_43 STEPHANE_ZANOTTI_F_49_7E	34	9 403
4	CFPP2000 07-02 LUCIE_DÉ_SILVIA_F_22_7E	47	11 593
5	CFPP2000 07-03 PAULINE_DÉ_BONNE_F_67_7E	96	19 634
6	CFPP2000 07-04 RAPHAËL_LAURENCE_H_23_7E	65	15 737
7	CFPP2000 07-05 YVETTE_AURÉAN_F_70_7E	100	22 040
8	CFPP2000 07-06 ROSEMONTE_EMMANUELO_FY_PATRICK_BERNARD_09_7E	67	16 940
9	CFPP2000 11-01 ANITA_MUSSO_F_46_11E	55	11 023
10	CFPP2000 11-02 MATTHÉO_LÉONORE_F_05 MAÏE LOUISE_OKSM_F_04_11E	110	29 180
11	CFPP2000 11-03 BATHOËLE_DICKHOF_H_05 JENNIFER_COURT_F_00_11E	102	27 684
12	CFPP2000 11-04 JULIE_TÉRIER_F_18 JULIA_TÉRIER_F_15_11E	81	20 974
13	CFPP2000 12-01 FÉLIX_BESSON_H_59 MAÏE_BESSON_F_X_12E	57	15 831
14	CFPP2000 12-02 ROSE_EMMANUELO_H_00 ROSE_DICKHOF_F_06_11E	78	19 083
15	CFPP2000 12-03 THÉRÈSE_LÉ_VENISE_F_70 VALENTIN_TÉRIER_F_00_12E	83	22 966
16	CFPP2000 12-04 MATTHÉO_ROSE_H_XV ÉLIE_FIGINI_F_XV_12E	94	23 126
17	CFPP2000 13-01 GABRIEL_PULCHER_H_40_13E	70	14 442
18	CFPP2000 14-01 NICOLÉ_NIKOVY_F_53_14E	88	22 282
19	CFPP2000 14-02 MIKA_F_08_14E	43	7 354
20	CFPP2000 18-01 PAUL_SMO_H_20 PIERRE_MILAN_SMO_H_34_18E	145	21 406
21	CFPP2000 20-01 GARY_COUGARE_H_24_20E	85	20 019
22	CFPP2000 20-01 JACQUELINE_FALUTIN_F_05_20E	87	17 331
23	CFPP2000 28-01 KILIAN_BELMONT_H_22 LUCAS_MERIANO_H_21_28E	65	12 756
24	CFPP2000 50-01 LOUISE_LOTARÉ_F_05 ET_JENNIFER_BILLET_F_75_50E	56	15 596
25	CFPP2000 50-02 YONKE_FERJAL_H_29 ARDEL_HIGRIM_H_25_50E	145	37 772
26	CFPP2000 80-01 ANNE_LINDING_H_58_80E	68	16 004
27	CFPP2000 80-02 MAÏE_HELENE_MATHÉA_F_57_80E	89	23 699
28	CFPP2000 90-01 DOMINIQUE_VALLÉE_F_37_90E	68	15 814
	TOTAL	2 198 <small>Soit 136,6 h</small>	517 774

Figure 1 : CFPP2000, index des entretiens

1.3. *Quoi* : exploration lexicométrique dans CFPP2000

Nous proposons ici une étude plus précise concernant *quoi*, forme qui sera décrite plus loin d'un point de vue sémantique, syntaxique et discursif. CFPP2000 contient 1086 occurrences de la forme graphique *quoi*. Le graphique suivant précise sa ventilation dans le corpus vu par le prisme des quartiers concernés :

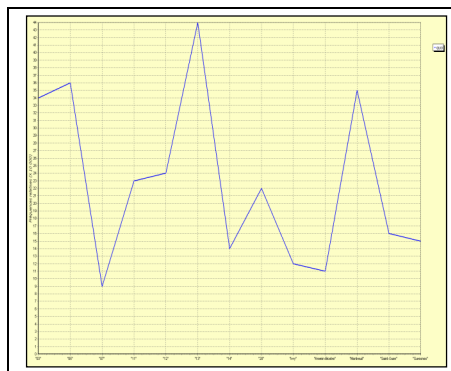


Figure 2 : *quoi*, ventilation par quartier

Quoi en contexte

La forme est utilisée de manière importante par quelques individus :

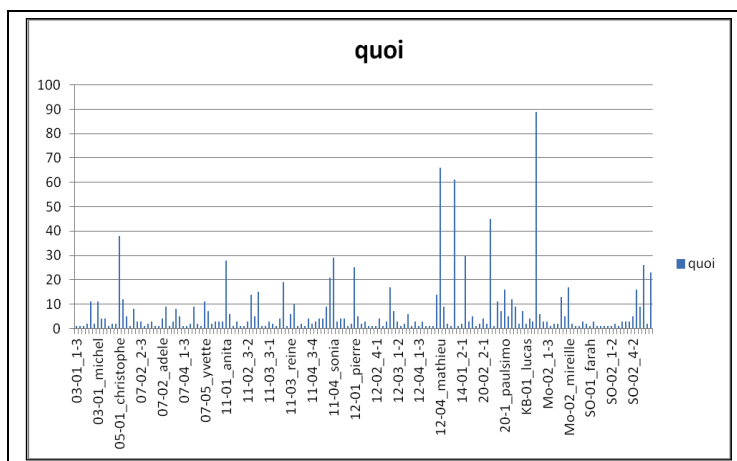


Figure 3 : quoi, ventilation par individu

Les individus qui l'utilisent plus de 10 fois dans l'ensemble de leurs tours de paroles sont regroupés dans le tableau suivant :

forme	03-01_1-3	03-01_michel	05-01_christophe	07-02_2-3	07-02_adele	07-04_1-3	07-05_yvette	11-01_anita	11-02_3-2	11-03_3-1	11-03_reine	11-04_3-4	11-04_sonia	12-01_pierre	12-02_4-1	12-03_1-2	12-04_1-3	12-04_mathieu	14-01_2-1	20-02_2-1	20-1_paulsimo	KB-01_lucas	Mo-02_1-3	Mo-02_mireille	SO-01_farah	SO-02_1-2	SO-02_4-2			
quoi	11	11	38	12	11	28	14	15	19	10	21	29	25	17	14	14	66	61	30	45	11	90	16	12	89	13	17	16	26	23

Figure 4 : répartition de quoi par individu

1.4. Au-delà de quoi

Réseau de cooccurents

La méthode des cooccurrences (Martinez, 2003) permet de mettre au jour le vocabulaire spécifique autour de la forme pivot sélectionnée. Le réseau résultant de cooccurents peut être interprété au regard des paramètres utilisés pour le calcul (Lafon, 1980; Martinez, 2003):

- **Fréquence** : fréquence de la forme dans le corpus
- **Co-Fréquence**: fréquence d'apparition conjointe dans un contexte donné du pivot et de son cooccurrent
- **Indice de spécificité** : indice indiquant la sur-représentation (ou la sous-représentation) d'une forme dans un contexte donné.
- **Nombre des contextes** : nombre de contextes dans lesquels apparaissent conjointement la forme pivot et son cooccurrent

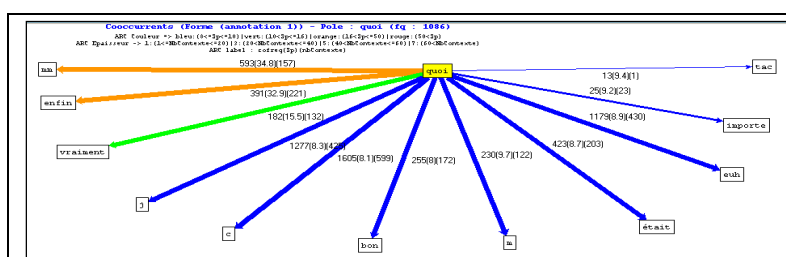
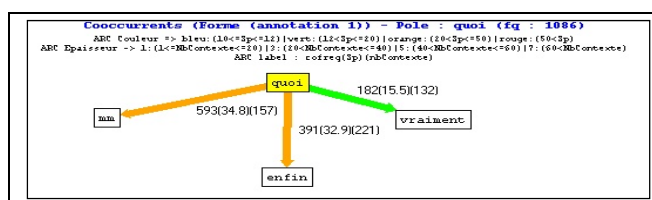
Dans les graphiques qui suivent, les arcs du réseau de cooccurrence portent une valeur regroupant ces 3 derniers paramètres : cofreq (indice de spécificité) (nombre de contexte). Les arcs du graphe se distinguent par leur couleur et leur épaisseur :

Couleur	Indice de spécificité	Épaisseur des arcs	Nombre de Contextes
Orange	13 ≤ 50	3	21-40

Vert	9 <= 12	5 ████████	41-60
Bleu	6 <= 8	7 ████████	60 et plus

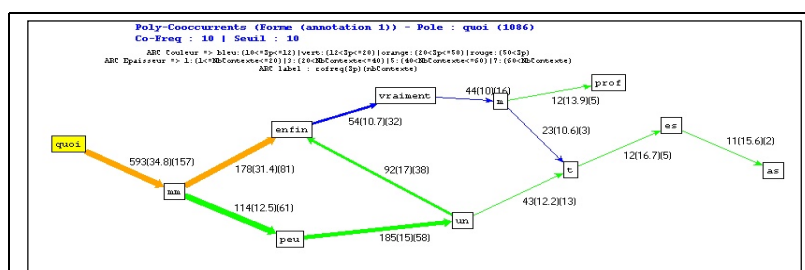
Figure 5 : Grille de lecture pour le réseau de cooccurrence

Les deux figures qui suivent correspondent aux 2 réseaux de cooccurents de la forme *quoi* associés aux paramètres de calcul indiqués :



Polycooccurrence

Le terme *poly-cooccurrence* désigne les attractions lexicales au-delà de la cooccurrence binaire. Le calcul de cette attraction reprend l’algorithme décrit dans Martinez 2006. Sur la forme *quoi*, ce calcul conduit au graphe suivant :



La grille de lecture de ce graphe est similaire à la précédente pour l’interprétation des valeurs numériques associées aux arcs ; le chemin d’attraction lexicale correspond au parcours du graphe dans le sens des arcs (*quoi*->*mm*->*enfin*->*vraiment*... ; *quoi*->*mm*->*peu*...).

Voyons à présent plus précisément l’étude du mot *quoi* dans CFPP2000 et en quoi ces calculs

lexicométriques peuvent enrichir cet examen.

2. Etude sémantique, syntaxique et discursive du mot *quoi*

Pour opérer l'analyse du mot *quoi* dans CFPP2000, nous prendrons pour cadre l'étude qui lui a été consacrée dans *Quoi de neuf sur quoi ? Etude morphosyntaxique du mot quoi* (Lefeuve, 2006) ainsi que les travaux qui ont amené à cette étude, principalement ceux de Le Goffic 2006. *Quoi* se caractérise par trois attributs.

i) C'est une variable en *qu-* refusant l'assertion :

**Il regarde quoi.*

ce qui le différencie d'un pronom tel que *quelque chose* :

Il regarde quelque chose.

La proforme *quoi* renvoie à des occurrences désignées par l'acte de prédication mais ne peut pas se fixer sur l'une d'entre elles ni sur plusieurs d'entre elles, même quelconques, comme si elle balayait tous les éléments d'une classe sans jamais se stabiliser sur un élément précis.

ii) *Quoi* renvoie à du non catégorisé

De même que les démonstratifs *ce, ça, cela, ceci* (cf. Kleiber 1994), *quoi* signifie le non catégorisé. Deux raisons à cela : il peut renvoyer à des humains :

Il donna un coup de pied au broc qui chancela, il fit claquer sa porte. « Elle m'a pris pour quoi ? Pour une midinette ? Je m'en fiche de sa dédicace... » (Sabatier, *Les Fillettes chantantes*, ex. tiré de Lefeuve 2006)

et à des structures prédicatives :

(35)[...] *les enfants entendirent la messe et déjeunèrent jusqu'à neuf heures ; après quoi, elles firent deux heures de lecture* (Chandernagor, *L'Allée du Roi*, ex. tiré de Lefeuve 2006)

C'est ce qui peut expliquer la combinaison aisée du pronom *quoi* avec ces démonstratifs :

mais c'était pas c'est ça à quoi j'pensais j'pensais aux oiseaux:::

CFPP2000 [11-02] Mathilde Lelong F 85 Marie Louise Orsin F 64 11^e

iii) Pour renvoyer à quelque chose qui n'est pas encore catégorisé ni classifié, le français a la particularité de recourir à deux pronoms interrogatifs, *que* et *quoi*. De nombreux linguistes ont montré que *que* et *quoi* adoptent un comportement syntaxique fort différent par rapport au verbe tensé. *Que* doit s'appuyer sur le verbe en restant joint à lui alors que *quoi* est plus autonome par rapport au verbe.

Ces trois paramètres — variable refusant l'emploi en assertion, sémantisme du non catégorisé, dualité *que* et *quoi* — permettent de comprendre les emplois syntaxiques de *quoi*. *Quoi* peut être interrogatif :

*ah y avait très peu d'passages de voitures on jouait à on jouait à **quoi** on jouait à des jeux où on s'cachait derrière les voitures*

CFPP2000 [03-01] Ozgur Kilic H 32 alii 3e

indéfini :

*euh c'est c'est vrai qu' ça démystifie l'truc clairement + + quand tu t'rends compte que de toutes façons **quoi** qu'tu fasses pour eux ils reviennent*

CFPP2000 [11-03] Blanche Duchemin F 25 Reine Ceret F 60 11e

*euh quand j'étais gamin y avait pas vraiment d'vie d'quartier c'est-à-dire que j'descendais d'chez moi je vois y avait une boulangerie XXX euh euh qui euh voilà j'y j'y allais euh elle me connaissait y avait des gros chiens enfin bon c'était + c'était une vie d'quartier **quoi***

CFPP2000 [03-01] Ozgur Kilic H 32 alii 3e

percontatif (interrogatif indirect) :

*quand vous m'avez parlé d' l'interview sur le septième j'me demandais + de **quoi** est-ce qu'on allait pouvoir parler pendant +*

CFPP2000 [07-04] Raphael_Lariviere_H_23_7e

intégratif (relatif sans antécédent) :

*C'est à quoi je pense euh mes jeunes collègues n'ont pas de **quoi** se payer ces appartements*

CFPP2000 [05-01] Christophe_Andre_H_62_Marie_Anne_Andre_F_63_5e

et relatif :

*mais c'était pas c'est ça à **quoi** j'pensais j'pensais aux oiseaux:::*

CFPP2000 [11-02] Mathilde_Lelong_F_85_Marie_Louise_Orsin_F_64_11^e

Dans CFPP2000, les 1086 occurrences de « *quoi* » se répartissent en deux emplois principaux, les emplois interrogatifs et l'emploi indéfini comme marqueur discursif. Le moteur de recherche permet de visualiser ces occurrences dans un contexte plus ou moins long.

En ce qui concerne les emplois interrogatifs, nous avons trouvé *quoi* dans des configurations qui tiennent compte des paramètres caractéristiques de *quoi*. La première comprend *quoi in situ* :

*ça fait **quoi** ?*

CFPP2000 [03-01] Ozgur_Kilic_H_32_alii_3e

Dans la deuxième configuration, *quoi* constitue le prédicat de l'énoncé qui correspond alors à une phrase averbale. Dans CFPP2000, on trouve 4 emplois avec *quoi d'autre* :

*moui l'entretien d' la maison oui les courses peu j'dirai + + et **quoi d'autre** ? non boh ++ oui entretien d' la maison oui + + oui mettre la table*

CFPP2000 [11-01] Anita_MUSSO_F_46_11^e

Pour ce qui est des emplois de *quoi* indéfini, il s'agit de *quoi* comme marqueur discursif (cf. Moignet 1981 et Lefeuvre 2006) où *quoi* se trouve en marge des énoncés assertifs, dans des emplois tels que :

*euh quand j'étais gamin y avait pas vraiment d'vie d'quartier c'est-à-dire que j'descendais d'chez moi je vois y avait une boulangerie XXX euh euh qui euh voilà j'y j'y allais euh elle me connaissait y avait des gros chiens enfin bon c'était + c'était une vie d'quartier **quoi***

CFPP2000 [03-01] Ozgur_Kilic_H_32_alii_3e

Selon nous (cf. Lefeuvre 2006 et Lefeuvre et al. 2011), *quoi* apparaît après un mot dont la formulation est problématique, ce qui peut expliquer la présence de ce *quoi* dans des contextes où l'hésitation est de mise, comme le montre par exemple la présence de marqueurs d'hésitation tels que *euh* ainsi que des lexèmes renvoyant explicitement au dire (*disons ci-dessous*) :

non bah euh ouais non musicien c'est sûr c'est pas + c'est pas un travail

[rires|noise|instantaneous] *non j'rigole c'est une blague c'est une boutade euh non non mais disons qu'c'est pas euh c'est pas un poste **quoi***

CFPP2000 [03-01] Ozgur_Kilic_H_32_alii_3e

La recherche de la bonne expression peut se traduire par la présence d'un paradigme qui comporte plusieurs segments comme autant de formulations possibles :

spk3 : *et puis après bah non euh: c'était toute la période:: c'est surtout Brigitte Bardot qui m'a frappée à c't'époque là et puis euh: Françoise Arnoux euh Charles Boyer des gens comme*

spk3 spk2 : ça **quoi** mmh mmh

spk3 spk1 : c'étaient mmh mmh

spk3 : les.: les x de

CFPP2000 [11-03] BLANCHE_DUCHEMIN_F_25_REINE_CERET_F_60_11E

(ex. tiré de Lefeuve et al. 2011)

Quoi se trouve régulièrement en fin d'unités rhématiques (cf. Morel et Danon-Boileau 1998 et Lefeuve et al. 2011) qui correspondent généralement à des unités syntaxiques organisées autour d'un prédicat :

*puis donc y avait un cour et c'est vrai qu'des fois l'week-end y avait euh quinze enfants avec des tracteurs à + à pédales des kartings enfin bon c'était euh c'était un un centre aéré **quoi** ++ mais non personne XXX*

CFPP2000 [03-01] Ozgur_Kilic_H_32_alii_3e

et qui peuvent clôturer un tour de parole :

*ils ont six ans et trois ans (mm) + et euh moi ça ça moi ça m'plaît beaucoup ça + d-d'avoir c'que j'ai c'qu'on a eu nous quand on était euh moi y avait qu'une majorité d'bobos euh mais euh de fils de de bobos mais euh beaucoup moins euh enfin y avait quand même + un mélange dans dans notre école **quoi***

CFPP2000 [03-01] Ozgur_Kilic_H_32_alii_3e

Il peut se combiner avec des marqueurs indiquant au contraire le début d'une unité syntaxique. C'est ce qu'indiquent les schémas de cooccurrence (cf. réseau de cooccurrents et polycooccurrence de la première partie) qui signalent l'affinité de *quoi* avec *enfin* ou *bon* :

*et ça moi je trouve ça très intéressant cette ambiance un peu internationale + des personnes des touristes qui vous demandent leur chemin à tous les coins de rue **enfin** ça fait vacances en même temps **quoi** je trouve ça vraiment sympa et euh ++ ben sinon c'est vrai qu'on a un cadre de vie vraiment **enfin** formidable on a + je sais pas*

CFPP2000 [07-01] Laurence_Leblond_F_43_Stephanie_Zanotti_F_49_7e

*ah non j' pense que papa et maman comme moi **bon** eux ils mangent du houmous **quoi** tu vois*

CFPP2000 [11-03] BLANCHE_DUCHEMIN_F_25_REINE_CERET_F_60_11E

Enfin marque un aboutissement de la réflexion en proposant une reformulation synthétique de ce qui précède ; *quoi* signale une recherche de formulation la plus proche possible du référent à exprimer :

*enfin je sais plus trop euh on avait pas la télé on avait pas **enfin bon** + c'était une vie ouvrière vraiment ouvrière ouvrière **quoi***

CFPP2000 [Mo-01] ANDRE_MORANGE_H_58_Mo

Bon quant à lui valide partiellement ce qui précède, ce qui incite le locuteur à ajouter une expression plus adéquate (cf. Lefeuve 2011).

Les schèmes de cooccurrence signalent également l'affinité de *quoi* avec *vraiment* :

*j'ai **vraiment** un super souvenir de ce : de c' quartier à cette époque **quoi** + plus la mixité euh + j'me rappelle alors là c'est beaucoup plus tard mais quand j'avais dix-huit ans à peu près + +*

CFPP2000 [11-01] Anita_MUSSO_F_46_11e

Vraiment est un modalisateur (cf. Kerbrat-Orecchioni 1980) qui permet au locuteur de présenter son dire comme adéquat au contenu de l'énoncé, d'où la recherche de la formulation qui corresponde le mieux à la réalité, recherche induite par le mot *quoi* et marquée ici par l'arrêt d'une première structure syntaxique au profit d'une deuxième structure :

mais c'est oui mais j'suis d'accord c'est une question sociale le + les classes sociales qui vivent au Kremlin-Bicêtre sont pas vraiment celles sont pas du tout celles avec lesquelles j'ai envie de de de m'épanouir plus tard quoi

(CFPP2000 [KB-01] Killian_Belamy_H_22_Lucas_Hermano_H_21_KB)

Le locuteur qui emploie le plus *quoi* (cf. première partie) utilise un patron syntaxique régulier en *c'est* + *vraiment* + *attribut* + *quoi* :

oui partout mais c'est pas Montreuil spécialement j pense que là c'est vraiment euh un phénomène plutôt général quoi

CFPP2000 [Mo-01] Andre_Morange_H_58_Mo

Quoi signale la recherche d'une expression et le fait que le segment proposé est le plus conforme à traduire la réalité, mais sans le valider. *Vraiment* valide le segment introduit. Nous avons donc plusieurs marqueurs qui signalent un lien avec la (re)formulation, que ce soit *vraiment* qui valide une formulation, *bon* qui valide partiellement ce qui précède, *enfin* qui synthétise ce qui précède et *quoi* qui marque la recherche de formulation.

3. Explorations sociolinguistiques avec le corpus CFPP2000

Le corpus CFPP2000 a été initialement conçu comme un outil permettant d'explorer le paysage linguistique de Paris, et notamment d'étayer des études variationnistes portant sur les parlars de la capitale française. Nous avons ainsi sélectionné des quartiers susceptibles de faire ressortir des différences entre le français standard actuel – généralement présenté comme une forme de français parisien « éduqué » (Fouché 1969, p. iii, Valdman 1982, p. 218) – et le français parisien « populaire », le « titi parisien », représenté dans un certain cinéma des années trente (Abecassis 2005). C'est ainsi que nous avons opposé les « beaux quartiers » du centre de Paris comme le 7^e arrondissement, à des communes relativement modestes de la petite ceinture comme Saint-Ouen ou Montreuil. Notre hypothèse de départ était toutefois celle d'un affaiblissement de ce parler régional au profit du standard, au niveau tant phonétique que lexico-grammatical, et nous avons effectivement constaté (sans pour l'instant avoir mené d'étude approfondie et statistique) une standardisation extrêmement dominante à toutes les tranches d'âge, avec tout au plus des survivances régionales chez certains locuteurs âgés, par exemple la réalisation du phonème *a* chez cette locutrice de 67 ans, habitant depuis son enfance à Montreuil:

des gâteaux des bons gâteaux du bon pain

CFPP2000 [Mo-02] MARIE_HELENE_MATERA_F_67_MO

Si la variable âge du locuteur nous permet de situer ce type de spécificité parisienne dans un cadre diachronique passé, d'autres exemples de variation non-standard peuvent relever d'innovations plus récentes et ainsi permettre de constater un changement en cours. C'est ce type de variation que nous allons explorer pour la forme *quoi*. Notre intérêt portera spécifiquement sur la dyade *c'est quoi* / *qu'est-ce que c'est*, lieu de variation reconnue même des locuteurs ordinaires. Ces deux formes se distinguent à plusieurs niveaux.

D'abord au niveau syntaxique il y a opposition entre une forme à inversion et une forme à ordre assertif. De manière générale, le français tend depuis longtemps vers l'interrogation sous l'ordre assertif (Frei 1929/2003, p. 349), la valeur interrogative étant portée par la seule prosodie. Cette tendance devrait en toute logique défavoriser la forme *qu'est-ce que c'est*, d'autant plus qu'il comporte deux « inversions » de type interrogatif : [*qu'*][*est-*][*ce*] : ordre interrogatif complément-verbe-sujet ; et [*qu'est-ce-*]-*que*-[*c'*][*est*], le premier syntagme

constituant un antécédent antéposé au sujet-verbe par le biais du relateur *que*. Mais cette complexité formelle est compensé par un figement syntagmatique prononcé (cf. **qu'était-ce que*). C'est ainsi que cette forme a pu se prémunir du destin de l'inversion interrogative.

La variation syntaxique en interrogation s'étend à trois formes, car aux ordres assertif et interrogatif s'ajoute une forme préfixale avec syntagme *est-ce que*. Ainsi, notre étude devrait porter sur trois variantes : *qu'est-ce / c'est quoi / qu'est-ce que c'est*. Mais le premier – l'ordre interrogatif simple **qu'est-ce* – n'est pas employé.

La forme à ordre assertif *c'est quoi* représente une flexibilité syntaxique accrue, l'élément visé par l'interrogative pouvant se placer avant ou après elle.

la finalité c'est **c'est quoi**

CFPP2000 [12-01] Pierre_Beysson_H_59_Marie_Beysson_F_X_12e

c'est une marque **c'est quoi** le maroual?

CFPP2000_SO-01_Louise_Liotard_F_85_et_Jeanne_Mallet_F_75_SO

Elle se prête ainsi parfaitement à la syntaxe cumulative du français parlé, la dislocation pouvant s'effectuer à droite ou à gauche. En revanche la forme *qu'est-ce que c'est* accepte difficilement, dans une interrogation allocutive, une antéposition du syntagme visé. La postposition, elle, peut se faire « nue » ou à l'aide du relateur *que*:

qu'est-ce que c'est que cet appareil? ++

CFPP2000_SO-01_Louise_Liotard_F_85_et_Jeanne_Mallet_F_75_SO

Enfin au niveau prosodique, l'accent tonique tombe sur le *quoi* de *c'est quoi* ; *qu'est-ce que c'est* renvoie celui-ci sur l'objet à définir. Cette saillance du *quoi*, anaphore d'un objet à définir, et donc symbole de l'information absente et devant être fournie par l'interlocuteur, peut conduire l'interlocuteur à ressentir la question comme une menace de face. La forme *qu'est-ce que c'est (que)* est plus longue (3 ou 4 syllabes au lieu de 2), sa distribution prosodique est moins polarisée, et elle accentue l'objet à définir, objet déjà présent dans le discours de l'autre.

La forme de l'interrogatif en *quoi* est souvent présentée comme étant une forme orale (« surtout fréquent dans la langue parlée », Grevisse, 1993, § 702). De la part des prescriptivistes elle n'est pas tant critiquée qu'ignorée. On trouve sur internet des reflets de cette surnorme, mais il faut sans doute se garder d'en exagérer la portée.⁶

Le corpus CFPP2000 recèle 80 occurrences de la variantes *ce + être + quoi*, pour 24 de *qu'est-ce que ce + être*⁷. Ces chiffres montrent donc une nette ascendance (76,9 %) de la forme à ordre assertif. Le résultat peut surprendre dans la situation de parole relativement formalisée de l'entretien (présence de l'enquêteur, personne inconnue ou distante).

Nous allons maintenant tenter de cerner le profil des utilisateurs de la forme minoritaire *qu'est-ce que ce + être*. Les 24 occurrences que contient le corpus sont le fait de 8 locuteurs seulement. Une locutrice – Sonia Branca-Rosoff – est à l'origine de 15 de ces occurrences

⁶ Voir par exemple la discussion du forum Français notre belle langue <<http://www.achyra.org/francais/viewtopic.php?p=60657>> (consulté 25 janvier 2012).

⁷ Sont exclues les occurrences comportant un *est* auxiliaire. Exemple: et alors euh l'autre chose c'est vous avez dit qu'votre père était artisan **qu'est-ce que c'est** devenu euh (Mo-01)

(62,5 % du total).

Cette dominance peut d'abord s'expliquer par la situation de parole. Un enquêteur est tout naturellement amené à *poser des questions*, on peut donc s'attendre à une fréquence inhabituelle de ces formes. Une autre enquêtrice, Florence Lefeuvre, est à l'origine de deux des énoncés. En comparant les chiffres d'emploi de la forme *ce + être + quoi* nous constatons que 51 des 80 occurrences sont également dues à l'enquêteur, une proportion similaire (voir tableau). Par ailleurs, si l'on accepte l'analyse d'une interrogative davantage « orale », voire critiquée, on peut trouver normal qu'il figure dans le discours d'une personne invitée à effectuer un enregistrement, le plus souvent au domicile familial.

	total	<i>enquêteur</i>	<i>non-enquêteur</i>
<i>ce + être + quoi</i>	80	51 (63,8 %)	29 (36,2 %)
<i>qu'est-ce que ce + être</i>	24	17 (70,8 %)	7 (29,2 %)

Les enquêteurs réalisent ainsi autour des deux-tiers des formes, la disparité étant plus marquée pour la forme en *qu'est-ce que*. Les enquêtés produisent en tout 36 occurrences des deux types, avec le type en *quoi* sélectionné dans 80,6% des cas. Il faut noter que dans la plupart des entretiens du corpus CFPP2000, un seul enquêteur était en face d'au moins deux interlocuteurs. Cette co-présence d'enquêtés a été adoptée afin de minimiser l'artificialité de la situation d'entretien et la distance entre les locuteurs; le fait peut avoir favorisé la forme moins soutenue *c'est quoi*. Par ailleurs, parmi les 7 énoncés « non-enquêteur » de *qu'est-ce que ce + être* 3 relèvent d'un discours direct *cité*, donc distancié de la situation de parole, « irréels » en quelque sorte. Ainsi la forme autocitationnelle suivante :

au marché j'ai dit "qu'est-ce que c'est que ça" !

CFPP2000 [MO-02] MARIE_HELENE_MATERA_F_67_MO

fait appel à l'imaginaire linguistique de la locutrice (les deux formes comparables qu'elle emploie lors de l'entretien sont de type *quoi*). Il en va de même pour le monologal

j'me dis "qu'est-ce que c'est cette chose ?"

CFPP2000 [SU-01] DOMINIQUE_VALIN_F_37_SU

ou le discours commun (pronom *on*) :

au début quand même l'APEIS était + plus ou moins reconnue parce qu'on disait "mais qu'est-ce que c'est que ceux-là euh"

CFPP2000 [IV-01] JACQUELINE_PELLETTIER_F_65_IVRY

Il ne reste ainsi en tout que 3 occurrences non-citationnelles de *qu'est-ce que ce + être* interrogatif dans le corpus, et 2 d'entre elles sont en séquence, la deuxième fortement conditionnée par la réalisation de la première. La moyenne d'âge des 8 locuteurs employant la forme *qu'est-ce que ce + être* est de 56 ans.

Pour la variation *qu'est-ce que c'est* vs *c'est quoi*, les 500 000 mots du corpus CFPP2000 confirment ainsi clairement la présence dominante de la forme *c'est quoi* dans le discours oral, fait posé chez Grevisse et ailleurs, mais il permet également d'identifier un contexte pragmatique-discursif – l'autocitation – favorisant l'une des formes, et de constater une surreprésentation de ces formes chez l'enquêteur. Ce dernier résultat porte sur le discours des

concepteurs de corpus eux-mêmes. Chez Sonia les interrogations en *qu'est-ce que* représentent une proportion significativement plus élevée du total des deux types (28,8 %), que chez les non-enquêteurs (19,4 %).⁸ Il est difficile de distinguer avec certitude l'influence de la position d'enquêtrice et celle de l'idiolecte, mais ce résultat « personnalisé » rappelle une étude menée par Sonia elle-même sur la variation entre *par contre* et *en revanche*, objet de mises en garde prescriptives; à l'issue de ce travail qui prenait appui sur le corpus CFPP2000 il est apparu que seules deux personnes utilisaient l'expression *en revanche*, une éditrice, Yvette Audin, et Sonia elle-même (Branca-Rosoff et al., 2009, p. 37-8).

Conclusion

Dans cette contribution nous avons tenté de montrer, à travers l'exemple du mot *quoi*, les exploitations disponibles et les résultats à attendre d'un grand corpus d'oral français outillé, le CFPP2000. Ce corpus, interrogeable librement sur internet, permet de disposer aisément de données pour l'exploration et la description du français oral. L'infrastructure proposée permet d'interroger des variables descriptives – âge, sexe, capital scolaire, etc. - à travers des métadonnées normalisées (Dublin Core, OLAC). La recherche d'occurrences par concordancier permet ensuite d'étayer une analyse lexico-grammaticale et discursive de mots ou de chaînes de mots spécifiques. Le croisement de ces deux sources d'information permet enfin de faire émerger des modélisations linguistiques, sous forme graphique ou statistique.

Bibliographie

- Abecassis Michaël (2005), *The representation of Parisian speech in the cinema of the 1930s*. Berne: Peter Lang.
- Branca Sonia, Fleury Serge, Lefeuvre Florence & Pires Mat (2009), « Constitution et exploitation d'un corpus de français parlé parisien », <http://cfpp2000.univ-paris3.fr/index.html>
- Chanet Catherine (2001), « 1700 occurrences de la particule *quoi* en français contemporain : approche de la « distribution » et des fonctions en discours », *Marges linguistiques*, n° 2, p. 56-80
- Delaveau Annie (2001), *Syntaxe, La phrase et la subordination*, Paris, Colin
- Fleury Serge (2007), *Le Métier Textométrique: Le Trameur, Manuel d'utilisation*. Université Sorbonne nouvelle Paris 3 (SYLED) <http://tal.univ-paris3.fr/trameur/>
- Fouché Pierre (1969), *Traité de prononciation française*. Paris: Klincksieck.
- Frei Henri (1929/ 2003), *La grammaire des fautes*. Rennes: Ennoïa.
- Grevisse Maurice (1993), *Le bon usage* (12e éd.). Louvain: Duculot.
- Haspelmath Martin (1997), *Indefinite pronouns*, Oxford, Oxford University Press.
- Kerbrat-Orecchioni Catherine (1980), *L'Enonciation, de la subjectivité dans le langage*, Paris : Armand Colin.
- Kleiber Georges (1994), *Anaphores et pronoms*, Louvain-la-Neuve, Duculot.
- Lafon P. (1980), Sur la variabilité de la fréquence des formes dans un corpus, *Mots*, 1 (octobre): 127-165.
- Le Goffic Pierre, 1993, *Grammaire de la phrase*, Paris, Hachette.
- Le Goffic Pierre (ed.), (2002), *Interrogation, indéfinition, subordination, Verbum*, n°XXIV, 4
- Le Goffic Pierre ed. (2006), *Les mots en Qu- du français, Lexique* n°18.

⁸ Type *qu'est-ce que*: Sonia Branca-Rosoff 15, non-enquêteurs 7; total *qu'est-ce que* + *quoi*: SBR 52, non-enquêteurs 36.

- Lebart L. and Salem, A. (1994), *Statistique textuelle*. Paris, Dunod, versions auteur disponibles en ligne : à l'ENST <http://ses.telecom-paristech.fr/lebart/ST.html> et à Paris 3 <http://www.cavi.univ-paris3.fr/lexicometrica/livre/st94/st94-tdm.html>
- Lefevre Florence (2011) « Bon dans le discours oral : une unité averbale autonome ? », in: Lefevre et Behr (eds), *Les énoncés averbaux autonomes entre grammaire et discours* (Paris, Ophrys) : 165-185.
- Lefevre Florence (2006), *Quoi de neuf sur quoi ? Etude morphosyntaxique du mot quoi*, Rennes : Presses universitaires de Rennes.
- Lefevre Florence, Morel Mary-Annick, Teston-Bonnard Sandra (2011), « Valeurs prototypiques de *quoi* à travers ses usages en français oral », *Neuphilologische Mitteilungen* (Bulletin de la Société Néophilologique, Helsinki) : 37-59
- Martinez W. (2003), *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*, Thèse pour le doctorat en Sciences du Langage, Université de la Sorbonne nouvelle - Paris 3.
- Martinez W., Leblanc J. -M. (2006), « L'analyse contrastive des réseaux de cooccurrence. Le 'monde' dans les discours des présidents de la Vème République ». Actes des 8èmes Journées d'Analyse Statistique des Données Textuelles, Univ. de Franche-Comté, Besançon, avril.
- Moignet Gérard (1981), *Systématique de la Langue française*, Paris, Klincksieck.
- Morel Mary-Annick et Danon-Boileau Laurent (1998), *Grammaire de l'intonation*, Paris, Ophrys.
- Valdman, A. (1982), « Français standard et français populaire: sociolectes ou fictions ? » *The French Review* 56 (2): 218-27.
- Vincent Diane (1993), *Les ponctuels de la langue et autres mots du discours*, Québec, Nuit Blanche.