

Différences fonctionnelles entre quasi-synonymes: les interrogatives de raison en français

Gabriel Thiberge*, Lisa Brunetti*, Jan Fliessbach+

*Université Paris Cité/CNRS +Potsdam Universität

1 Introduction

La présente étude examine les similitudes et les différences d'interprétation entre quatre expressions *wh* en français : *pourquoi*, *comment ça*, *comment se fait-il que*, et *qu'est-ce que* non-argumental, illustrées en (1).

- (1)
- a) Pourquoi tu as planté des oliviers dans ton jardin ?
 - b) Comment ça tu as planté des oliviers dans ton jardin ?
 - c) Comment se fait-il que tu as planté des oliviers dans ton jardin ?¹
 - d) Qu'est-ce que tu as planté des oliviers dans ton jardin ?

Ces quatre structures interrogatives ont été décrites dans la littérature comme compatibles avec une lecture de « raison », ainsi qu'avec d'autres lectures. Cette notion de « raison » a été appliquée à des expressions qui interrogent sur les conditions qui expliquent ou justifient un événement, mais aussi à d'autres qui concernent un événement considéré comme improbable ou impossible par la personne qui pose la question. Plus spécifiquement, *Pourquoi* a été décrit comme une expression qui demande une condition suffisante pour expliquer ou justifier un événement, mais aussi pour questionner sur le but (ou *purpose* ; Kédochim et Guesser, 2020), c'est-à-dire sur ce qui encourage un agent à agir. *Comment se fait-il que*, *comment ça* et *qu'est-ce que* non-argumental peuvent avoir une interprétation de « surprise » (Celle et Pélissier, 2022; Desmets et Gautier, 2009; Fleury, 2021; Fleury et Tovenà, 2018a, 2018b; Smirnova et Abeillé, 2021), à savoir demander les raisons de la possibilité d'un fait inattendu. *Qu'est-ce que* non-argumental questionnerait plutôt le but que la raison et aurait aussi une fonction de « reproche » (Dekhissi et Coveney, 2021). Des lectures de désapprobation, qui semblent étroitement liées aux reproches, ont également été attribuées à *pourquoi* (Celle et al., 2021; Myers et Pellet, 2014).

Jusqu'à présent, la littérature a créé un réseau de références croisées entre les lectures spécifiques de ces formes, sans toutefois en faire un compte rendu systématique. En effet, les similarités interprétatives (et les éventuelles différences) n'ont pas été explicitement présentées dans les travaux antérieurs. Dans cet article, nous tentons de combler ce manque en présentant les résultats de deux expériences en ligne, qui fournissent à notre connaissance la première comparaison expérimentale directe entre ces quatre formes et leurs significations. Nous faisons l'hypothèse que toutes les lectures mentionnées ci-dessus sont disponibles dans une certaine mesure pour chacune des interrogatives, mais qu'il existe des différences substantielles et quantifiables dans le degré d'association entre les quatre formes *wh* et les différentes lectures. Nos résultats montrent que ces interrogatives ne sont ni des variantes libres, à savoir ayant toutes la ou les mêmes fonctions, ni des formes en distribution complémentaire. Un modèle émerge de nos résultats qui est en partie conforme à la différenciation bien établie (Dayal, 2016; Farkas, 2022; Trotzke et Czymionka, 2022) entre questions pragmatiquement « canoniques » – à savoir ayant pour but de rechercher une information – et questions « non-canoniques », à savoir ayant une fonction expressive ou d'évaluation.

2 Études antérieures

Les études syntaxiques ont depuis longtemps remarqué la particularité des interrogatives introduites par des éléments *wh* correspondant à un adverbe de phrase (comme *pourquoi*) en français et dans d'autres langues. Un des pionniers de telles études, Korzen (1990), s'est posé la question « Pourquoi pourquoi est-il différent ? » à cause d'exemples comme (2) et (3), qui sont également abordées dans Myers et Pellet (2014).

- (2) N'importe [quand/*pourquoi].
- (3) ? *Pourquoi* pleure votre fils ?

Inspirée par Cornulier (1974), Korzen a noté que l'adverbe *pourquoi* diffère des autres formes interrogatives en ce qu'il est peu compatible avec l'inversion sujet-verbe et ressemble plutôt à *comment* « quand il s'emploie avec le sens : « comment se fait-il que », c'est-à-dire avec un sens causal » (Korzen, 1990, p. 61) (4).

- (4)
 - a) *Comment* Jean a-t-il survécu ?
 - b) ? *Comment* a survécu Jean ?
 - C'est qu'il est résistant. (Cornulier, 1974, p. 141)

En réalité, comme nous l'avons mentionné plus haut, une lecture causale ou de raison n'est qu'une parmi d'autres possibles pour *comment se fait-il que*. En fait, il semble qu'évoquer une lecture par le biais d'une autre forme *wh* est presque impossible, simplement parce que chacune d'entre elles a plusieurs lectures. Néanmoins, la similitude syntaxique entre *pourquoi* et une lecture spécifique de *comment* est remarquable. Comme le montre l'exemple (5), *comment ça*, *comment se fait-il que* et *qu'est-ce que* sont également problématiques dans des phrases avec inversion sujet-verbe.

- (5)
 - a) ? *Comment ça* pleure votre fils ?
 - b) ? *Comment se fait-il que* pleure votre fils ?
 - c) ? *Qu'est-ce que* pleure votre fils ?

Ces propriétés syntaxiques communes pourraient être liées au fait que ces éléments interrogatifs ne lient pas de variable syntaxique au sein de la phrase car ils interrogent sur des aspects de la proposition entière dénotée par la phrase, et pas sur une de ses parties (par exemple un argument du verbe ; Tovenà, 2023).

Les quatre éléments interrogatifs diffèrent dans leur forme, et trois sur quatre ont une forme marquée. *Comment ça* contient un démonstratif qui fonctionne comme une particule (Smirnova et Abeillé, 2021) ; *comment se fait-il que* dérive d'une phrase racine (ou indépendante, Abeillé et Godard, 2021, XXXVII) avec verbe *se faire* et sujet inversé impersonnel qui prend la préjacente comme complément ²; et finalement *qu'est-ce que* est une forme qui interroge sur l'objet direct mais qui devient non-argumentale quand elle a une lecture proche de celle de *pourquoi*. De plus, *qu'est-ce que* non-argumental est surtout utilisé avec des verbes qui portent une valeur expressive. Dekhissi et Coveney (2021, p. 134) ont principalement trouvé des occurrences avec *aller/venir*+infinitif ou avec des « verbes vernaculaires » ³. D'après les recherches limitées sur ces structures, il semble probable qu'au lieu de (1d), les locuteurs utiliseraient plutôt les interrogatives en (6). Néanmoins, le fait que ces chercheurs trouvent des exemples avec des verbes neutres (au plan expressif) nous permet une comparaison directe avec les autres formes *wh* qui peuvent être combinées avec tous les verbes ⁴.

- (6) a) Qu'est-ce que t'avais besoin de planter des oliviers dans ton jardin ?
- b) Qu'est-ce que tu viens te planter des oliviers dans ton jardin ?

Étant donné ces propriétés, et en supposant une correspondance entre propriétés non-canoniques de forme et de fonction (Farkas, 2022), on peut s'attendre à ce que les trois formes marquées déclenchent une interprétation non-canonique. Ceci est vrai pour *comment ça*, qui est associé à trois lectures dans Fliessbach et al. (under review) : demande de clarification, surprise, et désaccord, tandis que les deux dernières ne sont qu'en partie ou pas du tout des fonctions questionnantes.

Pour ce qui est de *qu'est-ce que* non-argumental, la littérature n'est pas concluante. Dekhissi et Coveney (2021, p. 131) affirment que ces interrogatives ressemblent à celles en anglais introduites par *what for*, en ce qu'elles sont « plus expressives que leurs équivalentes dans la langue standard et elles sont toutes deux exclues d'un contexte négatif, apparemment parce qu'elles portent sur l'objectif plutôt que sur la raison ». Cette affirmation doit être vérifiée : les réponses à une enquête préliminaire que nous avons menée – ainsi que nos propres intuitions – montreraient que *qu'est-ce que* est acceptable avec négation dans certains contextes, comme (7).

(7) Qu'est-ce que t'as même pas été lui dire bonjour ?

Ceci devrait cependant être confirmé par une étude expérimentale rigoureuse. Ces interrogatives ont quoi qu'il en soit aussi été décrites comme « conflictuelles », toujours par Dekhissi et Coveney (2021), en ce qu'elles expriment « l'agacement, voire l'indignation, du locuteur à l'égard des paroles ou des actions de son interlocuteur sous la forme d'une assertion emphatique implicite ». (Dekhissi et Coveney, 2021, p. 129) D'ailleurs, Myers et Pellet (2014, p. 169) ont montré que même dans le cas de *pourquoi*, la fonction de demande d'explication sert souvent à indexer un large éventail d'actes de langage : « l'expression de la surprise, du désaccord ou de la désapprobation, jusqu'à la remise en cause du point de vue de l'interlocuteur précédent ou des actions de l'un des interlocuteurs »⁵. Kédochim et Guessier (2020) ont montré que *pourquoi* ne demande pas seulement la raison, mais peut aussi demander le but. La différence est subtile : suivant Tsai (2008), ces auteurs soutiennent que l'interprétation de raison impliquerait une relation appelée *Enable* 'permettre', selon laquelle un événement est une condition nécessaire pour un autre, alors que l'interprétation de but impliquerait une relation *Motivate* 'motiver', dans laquelle un événement rend possible ou cause un autre, par l'intermédiaire d'un sujet agent.

La similarité d'interprétation entre ces différentes formes interrogatives est ponctuellement mentionnée dans la littérature. Pour l'exemple (8), Abeillé et Godard (2021, p. 1114) constatent que « la tournure figée *comment se fait-il que* signifie 'pourquoi' ».

(8) Comment se fait-il que les épisodes de certaines séries télévisées sont diffusés dans le désordre ? (csa.fr, 7 nov. 2012)

D'autre part, ce rapprochement reste anecdotique et aucune recherche antérieure ne porte, à notre connaissance, sur une comparaison interprétative systématique entre ces quatre formes. L'objectif de notre travail est ainsi d'aborder une telle comparaison en testant expérimentalement les intuitions des locuteurs sur la possibilité des quatre interrogatives d'avoir des lectures de raison, but, surprise, et reproche. Nous allons décrire en détail ce travail expérimental dans la section suivante.

3 Étude expérimentale

Pour explorer les interprétations principales des quatre expressions interrogatives *pourquoi*, *comment ça*, *comment se fait-il que* et *qu'est-ce que*, nous avons construit deux expériences, sur une instance hébergée sur des serveurs universitaires de la plateforme IbexFarm (Drummond, 2016). Ces deux études avaient donc deux motivations principales. D'abord, elles constituent, à notre connaissance, les premiers travaux expérimentaux de cette ampleur sur cette thématique : en cela, elles visent à observer de façon plus systématique les sens associées à ces formes interrogatives (nombre important de volontaires, plusieurs observations par personne et par condition par le recours au carré latin et à des listes contrebalancées,

analyses statistiques appropriées). Ensuite, elles sont aussi – toujours à notre connaissance – les premiers travaux à contraster *entre elles* les différentes interprétations associées à ces formes, en les soumettant, dans la même tâche, aux volontaires (à nouveau, en appliquant les principes clefs de la méthodologie psycholinguistique).

Chaque expérience comportait une phase d'information présentant les droits des volontaires, une phase d'instructions décrivant la tâche à faire, une phase d'entraînement, puis la tâche expérimentale à proprement parler, avant une validation de la passation.

Nous avons analysé les données de ces expériences avec le langage de programmation R et l'interface logicielle Rstudio (Posit team, 2023; R Core Team, 2023) dans le cadre de l'analyse statistique inférentielle bayésienne. En particulier, ces analyses permettent une réflexion moins binaire que le cadre fréquentiste classique : cela signifie qu'il n'y a pas d'opposition entre effets significatifs et non-significatifs, ce qui est plus approprié pour les études exploratoires comme la nôtre (voir Sorensen et al., 2016 pour plus de détails). Plus d'explications sur ce cadre d'analyse et sa grille de lecture, ainsi que les résultats détaillés de ces expériences (modèles, scripts, etc.), sont disponibles dans le répertoire en ligne https://osf.io/swn48/?view_only=d4213d64e12e41f686673b584bd097d7.

3.1 Expérience 1

La première étude visait à explorer l'interprétation des quatre formes interrogatives de façon assez indirecte. Nous avons construit des séquences énonciatives (deux phrases, d'abord une interrogative introduite par une des quatre formes, puis une séquence de type varié forçant une interprétation particulière de cette interrogative, au sein d'un très court contexte). Les volontaires devaient juger la cohérence de la continuation fournie avec l'interrogative présentée. L'idée était ici qu'une cohérence jugée globalement basse par les volontaires traduirait l'incompatibilité entre une interrogative donnée et une ou des continuations parmi les quatre envisagées (évoquant des lectures de but, raison, surprise, reproche). La comparaison de toutes les conditions expérimentales entre elles devait permettre de potentiellement faire apparaître : 1^o, une hiérarchisation des interprétations pour chacune des formes interrogatives étudiées (« la forme interrogative X est plutôt interprétée comme interrogeant sur Y ») et, 2^o, des associations privilégiées entre forme(s) interrogative(s) et interprétation(s) (« pour interroger sur X en français, il vaut mieux utiliser la forme Y »).

3.1.1 Protocole expérimental et volontaires

80 adultes L1 parlant le français (recrutement sur la plateforme Prolific, <https://www.prolific.com/>) ont jugé la cohérence entre l'une des quatre phrases interrogatives et une phrase suivante du même locuteur, manipulée pour favoriser l'une des quatre interprétations comme dans (9).

- (9) Pourquoi tu as planté des oliviers dans ton jardin ? ...
- a) Ils étaient en promotion ? (raison)
 - b) C'est pour produire de l'huile d'olive ? (but)
 - c) On devrait planter des fleurs pour aider les abeilles ! (reproche)
 - d) Tu as toujours parlé d'acheter des pommiers ! (surprise)

La variable dépendante était le jugement de cohérence prononcé par les volontaires (101 niveaux à la graduation non-perceptible sur une échelle avec un curseur à faire glisser avec la souris d'ordinateur, de *pas du tout cohérente* à *très cohérente*, recodés entre 0 et 100). Une première variable indépendante était le type de forme interrogative (4 niveaux ; *pourquoi*, *comment ça*, *comment se fait-il*, *qu'est-ce que* ; *pourquoi* en tant que niveau de référence). La seconde variable indépendante était la continuation (4 niveaux ; *raison*,

but, surprise, reproche ; raison en tant que niveau de référence). Ces $4 \times 4 = 16$ conditions ont été présentées de façon équilibrée et contrebalancée aux volontaires par le biais de plusieurs listes, où étaient mélangés 16 items cibles (voir Annexe OSF) et 16 items distracteurs. $16 \times 80 = 1280$ observations ont été recueillies sur les items cibles.

Pour les items cibles nous avons choisi l'ordre sujet-verbe car nous considérons qu'il est plus caractéristique du français parlé, en particulier dans les variétés « non standard » et dans les registres qui se prêtent à des actes de langage expressifs (Hamlaoui, 2011; Lambrecht, 1981; Massot et Rowlett, 2013). Les phrases contenaient toutes un verbe au passé composé et aucun de ces verbes n'appartenait pas aux catégories qui, chez Dekhissi et Coveney (2021), favorisent la présence de *qu'est-ce que*, et qui peuvent être considérés comme intrinsèquement expressifs.

3.1.2 Résultats

Comme l'illustre les Figures 1 et 2 (deux visualisations des mêmes données), *qu'est-ce que* a été jugé moins cohérent que les trois autres formes interrogatives pour tous les types de continuation, ce qui suggère un facteur externe (cf. *infra*). Les trois autres formes ont reçu des scores de cohérence élevés dans l'ensemble, démontrant un certain degré de similarité fonctionnelle. Par ailleurs, les continuations favorisant une lecture de raison ont été jugées moins cohérentes dans l'ensemble. Nous attribuons ce résultat au cadre expérimental, car il n'est typiquement pas nécessaire d'avoir une continuation après des interrogatives avec la lecture « non marquée » de raison. En effet, non seulement il s'agit d'une lecture questionnante, donc le locuteur est supposé céder le tour et attendre une réponse de son interlocuteur, mais elle est la lecture par défaut, on ne s'attend donc pas qu'elle nécessite une continuation qui la désambiguïse.

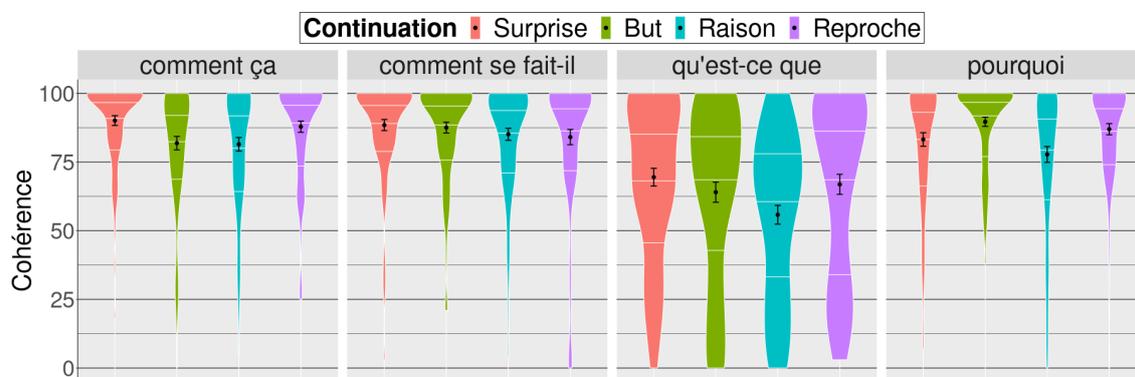


Figure 1. Cohérence des séquences en fonction de l'interrogatif et des continuations (Exp. 1)

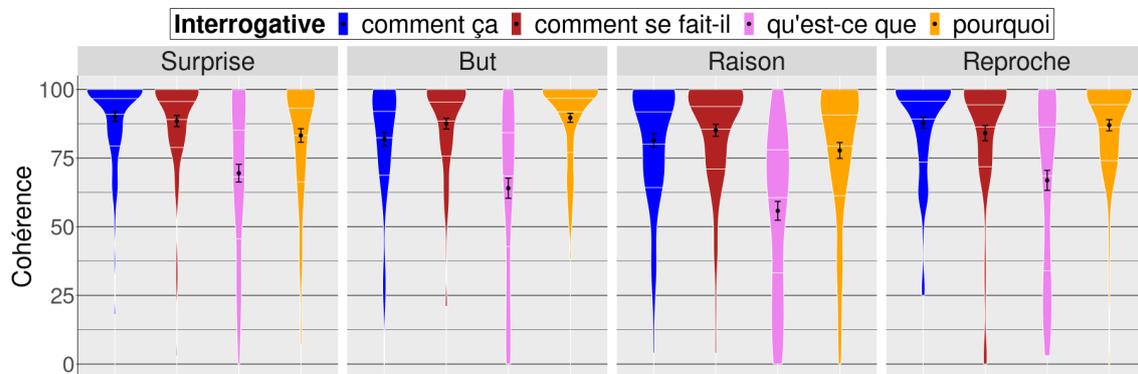


Figure 2. Cohérence des séquences en fonction des continuations et de l'interrogatif (Exp. 1)

Un modèle de régression linéaire bayésienne a été calculé avec les évaluations sur l'échelle de cohérence (semi-continue) comme variable dépendante, et le type d'interrogatif (4 niveaux) et la continuation (4 niveaux) comme variables indépendantes (toutes deux centrées autour de la moyenne pour une meilleure interprétation des effets ; Brehm et Alday, 2022). Avec Featherston (2021, p. 49), nous supposons que la dimension psychologique sous-jacente appliquée par les participants est une échelle d'intervalle. L'utilisation d'un curseur à 100 points a en outre facilité l'utilisation d'un modèle métrique (Sullivan et Artino, 2013, p. 541) car il ne court pas les risques associés aux échelles de Likert à cet égard (Liddell et Kruschke, 2018) : les volontaires interprètent l'échelle comme un continuum et dans l'analyse la variable dépendante est considérée comme continue et non-discrète. Les variables aléatoires comprenaient les scènes (groupes de quatre items avec la même préjacente mais pas le même élément interrogatif ni la même continuation, voir Annexe OSF) et les volontaires, avec des pentes aléatoires appropriées. En prenant *pourquoi* et *raison* comme niveaux de référence, nous avons trouvé plusieurs effets notables ⁶ :

a) des évaluations plus basses pour la forme *comment ça* avec des continuations de but (Coef.Est. = -8.24, $P(\text{Coef.} < 0) = .92$, 95%CrI = [-21.57; 4.51]) ;

b) des évaluations plus basses pour la forme *comment se fait-il que* avec des continuations de reproche (Coef.Est. = -7.86, $P(\text{Est.} < 0) = .90$, 95%CrI = [-21.49; 5.25]) et de but (Coef.Est. = -5.06, $P(\text{Est.} < 0) = .86$, 95%CrI = [-15.20; 4.69]) ;

c) des évaluations plus basses pour la forme *qu'est-ce que* avec des continuations de but (Coef.Est. = -4.91, $P(\text{Est.} < 0) = .80$, 95%CrI = [-17.91; 7.69]) ;

d) des évaluations plus hautes pour la forme *qu'est-ce que* avec des continuations de surprise (Coef.Est. = 8.27, $P(\text{Est.} > 0) = .90$, 95%CrI = [-6.63; 22.82]).

Pris dans leur globalité, ces résultats nourrissent deux perspectives sur la relation entre formes interrogatives et interprétations. D'abord, du point de vue des formes, certaines d'entre elles semblent plus cohérentes avec certaines continuations, ce qui montrerait qu'elles sont spécialisées pour exprimer certaines interprétations. Par exemple, *qu'est-ce que* et *comment ça* sont meilleurs avec des continuations de reproche et de surprise qu'avec des continuations de raison ou de but : cela montrerait que ces formes tendent à se spécialiser dans une lecture non-canonique de type expressif, tandis que *pourquoi* semble préférer une lecture questionnante, sur la raison ou le but. Par ailleurs, du point de vue des interprétations, certaines d'entre elles semblent favoriser le recours à certaines formes : par exemple, l'expression de la surprise semble plus cohérente avec *comment ça* ou *comment se fait-il que*, l'expression du reproche avec *comment ça*, l'expression du but avec *comment se fait-il que* et *pourquoi*, et enfin l'expression de la raison avec *comment se fait-il que*.

Concernant une possible influence de la variation sociolectale dans les jugements de cohérence pour les quatre formes interrogatives, une approche globale des données montre que les jugements des volontaires sont plus dispersés pour *qu'est-ce que* que pour les trois autres formes (Figures 1 et 2). Pour aller plus en détail, nous avons éliminé la variance intra-individuelle de l'analyse en agrégeant les observations par volontaire par condition. Une comparaison des erreurs autour de la moyenne dans ce jeu de données met en lumière que les jugements présentent toujours une dispersion près de deux fois plus importante autour de la moyenne pour *qu'est-ce que* ($SE_{question}=11.10$)⁷ que pour les trois autres formes ($SE_{commentsefaitilque}=6.75, SE_{commentça}=6.60, SE_{pourquoi}=7.50$). Au-delà de l'association de *qu'est-ce que* avec les continuations de surprise et de reproche, cette importante variabilité inter-individuelle indique que les évaluations généralement basses de cette forme sont au moins en partie d'origine sociolinguistique.

3.2 Expérience 2

L'expérience 1 explorait les représentations intuitives des volontaires sur les liens entre formes interrogatives et interprétations possibles. L'expérience 2 a été conçue comme un moyen d'explorer explicitement ces représentations en proposant aux volontaires les quatre interprétations possibles en même temps, de manière à ce qu'ils et elles puissent réfléchir au « poids » relatif de chacune des interprétations sur un même item. Une telle méthodologie est déjà utilisée par exemple (Fontaine et al., 2013; Kohler, 2005; Osgood, 1952), mais aussi dans les études en sociolinguistique expérimentale où plusieurs échelles de « jugement social » sont présentées en même temps aux volontaires (Kircher, 2016; Lambert et al., 1960). Cette méthodologie est plus métalinguistique que celle de l'expérience 1. Nous considérons que cela est avantageux pour la triangulation des différentes interprétations.

En effet, comme évoqué plus haut, la tâche de l'expérience 1 pourrait avoir favorisé les lectures non-canoniques à cause de la présence d'une continuation (de la part du même locuteur). Cette continuation est en soi inattendue si on suppose que la fonction par défaut d'une interrogative est le questionnement (Dayal, 2016, p. 2).

De plus, les scores globalement bas de *qu'est-ce que* pourraient indiquer que certaines personnes ont interprété l'Expérience 1 comme une tâche de jugement d'acceptabilité (au moins indirectement), et que ces personnes ont jugé négativement une expression diatopiquement et diastratiquement restreinte comme *qu'est-ce que*, indépendamment des continuations. De manière analogue, les scores élevés de *comment se fait-il que* peuvent en partie être attribués à son registre formel (du fait de l'inversion entre sujet et verbe ; Ashby, 1977; Quillard, 2001; Thiberge, 2020). La nature explicitement sémantique de la tâche de cette nouvelle expérience visait à limiter les effets sociolinguistiques présents dans l'expérience 1.

3.2.1 Protocole et volontaires

80 adultes L1 parlant le français (recrutement sur Prolific en excluant les personnes ayant participé à la première expérience), ont explicitement évalué les intentions d'une personne ayant produit une phrase interrogative, en donnant leur avis sur cinq échelles à curseur en parallèle (101 niveaux imperceptibles, recodés entre 0 et 100), à savoir les quatre interprétations sondées dans l'Expérience 1 (*raison, but, surprise, reproche*) auxquelles a été ajoutée *persuasion*, une échelle distracteur correspondant à l'interprétation probable de certains des items distracteurs ajoutés. Une description précise de ce que chaque échelle visait était présentée en parallèle des instructions données au début de l'expérience. Cette fois-ci, nous n'avons pas ajouté de phrases de continuation et seule l'interprétation de l'interrogative était donc visée. Pour limiter les biais potentiels, nous avons équilibré l'ordre de présentation des échelles avec différentes listes, chacune des listes ayant le même design global : 4 conditions présentées en carré latin, 16 items cibles, 32 items distracteurs, pour un total de 1280 observations par échelle sur les items cibles.

3.2.2 Résultats

Les Figures 3a et 3b offrent deux perspectives sur les mêmes résultats, avec une présentation simultanée des valeurs données par les volontaires pour, simultanément, chacune des quatre formes interrogatives étudiées et chacune des quatre échelles cibles de réponse (but, raison, surprise, reproche).

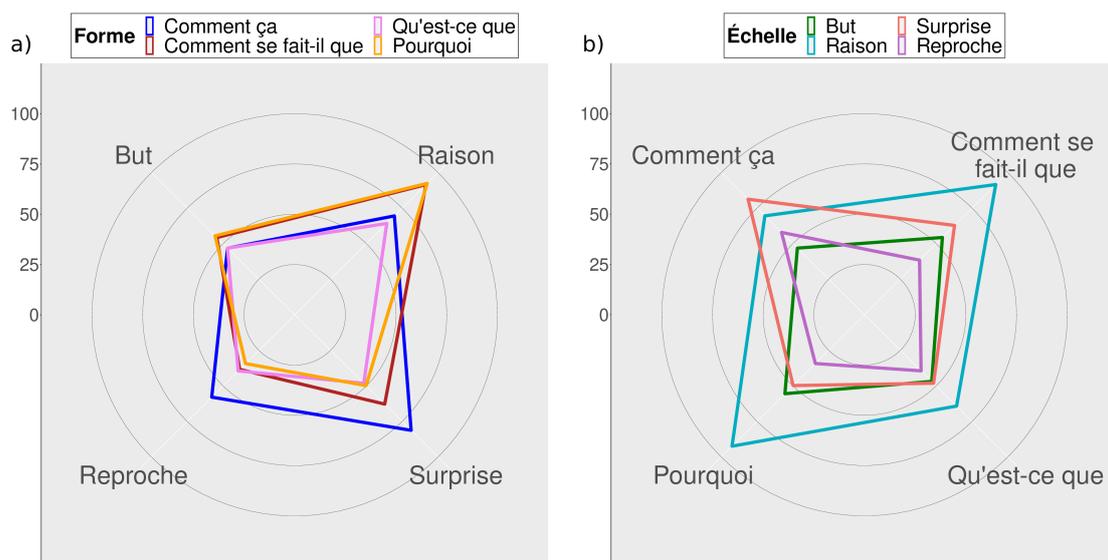


Figure 3 a,b. Jugements des intentions en fonction de l'interrogatif employé (Exp. 2)

Une matrice de corrélation complète est présente dans l'Annexe OSF et permet d'évaluer les liens créés entre les différentes échelles par les volontaires, avec une corrélation notable – mais attendue – par exemple entre l'échelle de surprise et celle de reproche (ρ de Spearman = 0.45, $p < 0.01$). Ces liens entre échelles ne paraissent dans l'ensemble pas incompatibles avec une étude en isolation de chacune, et en conséquence un modèle de régression linéaire bayésien a été calculé de façon autonome pour chacune des échelles de réponse. Pour chacun de ces modèles, la variable dépendante était la valeur donnée par les volontaires sur l'échelle en question (variable semi-continue, recodée entre 0 et 100), et la variable indépendante était la forme interrogative (quatre niveaux, avec *pourquoi* en référence). Les effets aléatoires comprenaient les volontaires et les items, avec des pentes aléatoires appropriées.

En prenant *pourquoi* comme niveau de référence pour chacune, nous avons trouvé plusieurs effets notables (liste plus complète dans l'Annexe OSF) :

- des interprétations de raison plus basses pour les formes *comment ça* (Coef.Est. = -22.17, $P(\text{Est.} < 0) = 1$, 95%CrI = [-29.50; -14.80]) et *qu'est-ce que* (Coef.Est. = -27.78, $P(\text{Est.} < 0) = 1$, 95%CrI = [-35.33; -20.23]) ;
- des interprétations de but plus basses pour les formes *comment ça* (Coef.Est. = -6.82, $P(\text{Est.} < 0) = .99$, 95%CrI = [-12.12; -1.64]) et *qu'est-ce que* (Coef.Est. = -7.64, $P(\text{Est.} < 0) = .99$, 95%CrI = [-13.62; -1.62]) ;
- des Interprétations de surprise plus hautes pour les formes *comment ça* (Coef.Est. = 31.45, $P(\text{Est.} > 0) = 1$, 95%CrI = [26.14; 36.78]) et *comment se fait-il que* (Coef.Est. = 12.71, $P(\text{Est.} > 0) = 1$, 95%CrI = [8.52; 16.82]) ;
- des interprétations de reproche plus hautes pour la forme *comment ça* (Coef.Est. = 20.86, $P(\text{Est.} > 0) = 1$, 95%CrI = [15.30; 26.40]) et *qu'est-ce que* (Coef.Est. = 4.04, $P(\text{Est.} < 0) = .86$, 95%CrI = [-3.44; 11.63]).

Pris dans leur ensemble, ces résultats complètent ceux de la première expérience. De manière générale l'interprétation de raison semble rester présente derrière les formes *comment ça* et *qu'est-ce que*, même si

elle favorise plutôt le recours à *comment se fait-il que* et *pourquoi* (lesquels sont aussi privilégiés pour l'interprétation de but). *Comment ça* semble bien véhiculer une interprétation dominante de surprise, et à un moindre niveau de reproche, ce qui apparaît ici plus nuancé pour *qu'est-ce que* (possiblement à nouveau en raison de son caractère sociolectal). À l'inverse, *comment se fait-il que* et *pourquoi* semblent bien parmi les formes moins compatibles avec une interprétation de type expressif, à savoir de reproche ou de surprise.

4 Discussion générale et conclusions

Les résultats de nos expériences corroborent plusieurs des nombreuses intuitions énoncées plus haut et issues de publications précédentes. Elles fournissent toutefois un cadre bien plus clair et plus précis des similitudes et des différences d'interprétation entre les quatre formes interrogatives étudiées.

Nous avons remarqué que la présence d'une continuation dans l'Expérience 1 a pu réduire la différence entre *pourquoi* et les autres formes, en rendant la valeur questionnante moins disponible. Comme nous l'avons dit plus haut, la présence de la continuation a facilité une lecture non-canonique, même avec *pourquoi*. Nous avons prédit que cela ne devrait pas être le cas dans l'Expérience 2. En effet, sans continuation, nous avons confirmé que *pourquoi* est l'élément le plus approprié pour questionner la raison et, plus généralement, qu'il est moins bon/approprié que les autres formes dans des usages non liés à la recherche d'information (surprise et reproche) (voir Figure 3 b). Les résultats montrent aussi que *comment ça* et *qu'est-ce que* sont moins appropriés que *pourquoi* pour demander un but, confirmant ainsi encore une fois leur utilisation moins optimale en tant que véritables questions. Les résultats sur *comment ça* montrent particulièrement bien sa préférence pour des usages expressifs. Il faut ici rappeler que la fonction de question de clarification (Smirnova et Abeillé, 2021) n'a pas été prise en compte dans les deux expériences, et qu'elle n'a donc pas été testée.

Cependant, nous n'avons trouvé aucune différence notable entre *comment se fait-il que* et *pourquoi* dans l'association avec la raison, ce qui est inattendu vu la forme marquée de *comment se fait-il que*. L'utilisation de cette forme dans un registre soutenu pourrait expliquer son rapprochement avec *pourquoi*. Une future comparaison avec le plus familier *comment ça se fait que* en dira sans doute plus sur sa place dans l'espace d'usage des interrogatives « de raison ».

Qu'est-ce que a été très mal jugé par les volontaires et donc les résultats à son propos devront être prolongés par d'autres travaux. Une analyse plus fine des résultats pourrait indiquer la présence de sociolectes qui autoriseraient plus ou moins bien son usage. Il serait donc utile dans une étude à part de sélectionner des personnes qui acceptent cette forme, et de ne tester ses fonctions que sur cette population.

Les différents protocoles expérimentaux des deux expériences nous ont aussi révélé l'importance, pour l'interprétation des interrogatives, du contexte interactionnel dans lequel elles se produisent. L'expérience 1 teste un seul tour de parole alors que les lectures questionnantes auraient plus de visibilité si le tour changeait et s'il y avait une réponse de la part d'une autre personne. De futures études ciblant davantage de données conversationnelles pourraient bénéficier de la prise en compte du contexte interactionnel dans le protocole expérimental.

Références

- Abeillé, A. et Godard, D. (2021). *La grande grammaire du français*. Actes sud.
Ashby, W. J. (1977). Interrogative forms in Parisian French. *Semasia*, 4, 35–52.
Brehm, L. et Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125, 104334. <https://doi.org/10.1016/j.jml.2022.104334>

- Brunetti, L., Tovenà, L. M. et Yoo, H. (2022). French questions alternating between a reason and a manner interpretation. *Linguistics Vanguard*, 8(s2), 227–237. <https://doi.org/10.1515/lingvan-2020-0130>
- Brunetti, L., Yoo, H., Tovenà, L. et Albar, R. (2021). French reason-comment ('how') questions: A view from prosody. Dans A. Trotzke et X. Villalba (dir.), *Expressive Meaning Across Linguistic Levels and Frameworks*. Oxford University Press.
- Celle, A., Jugnet, A. et Lansari, L. (2021). Expressive questions in English and French : *What the hell* versus *Mais qu'est-ce que*. Dans A. Trotzke et X. Villalba (dir.), *Expressive Meaning Across Linguistic Levels and Frameworks* (p. 138–166). Oxford University Press. <https://doi.org/10.1093/oso/9780198871217.003.0008>
- Celle, A. et Péliissier, M. (2022). Surprise questions in spoken French. *Linguistics Vanguard*, 8(s2), 287–302. <https://doi.org/10.1515/lingvan-2020-0109>
- Cornulier, B. de. (1974). "Pourquoi" et l'inversion du sujet non clitique. Dans C. Rohrer et N. Ruwet (dir.), *Linguistische Arbeiten : Vol. 13. Études de syntaxe : Actes du Colloque Franco-Allemand de Grammaire Transformationnelle* (p. 139–163). de Gruyter. <https://doi.org/10.1515/9783111611761-010>
- Dayal, V. (2016). *Questions* (First edition). *Oxford surveys in semantics and pragmatics : Vol. 4*. Oxford University Press.
- Dekhissi, L. et Coveney, A. (2021). Le contexte linguistique des questions rhétoriques conflictuelles et la variation entre *pourquoi* et *qu'est-ce que*. Dans P. Larrivée et A. Guryev (dir.), *Langue française : Vol. 212. Variantes formelles de l'interrogation* (p. 123–137). Armand Colin.
- Desmets, M. et Gautier, A. (2009). « Comment n'y ai-je pas songé plus tôt ? », questions rhétoriques en comment. *Travaux De Linguistique*, 58(1), 107–125. <https://doi.org/10.3917/tl.058.0107>
- Drummond, A. (2016). *IbexFarm (Version 0.3. 9)*.
- Farkas, D. F. (2022). Non-Intrusive Questions as a Special Type of Non-Canonical Questions. *Journal of Semantics*, 39(2), 295–337. <https://doi.org/10.1093/jos/ffac001>
- Featherston, S. (2021). Response Methods in Acceptability Experiments. Dans G. Goodall (dir.), *The Cambridge Handbook of Experimental Syntax* (p. 39–61). Cambridge University Press. <https://doi.org/10.1017/9781108569620.003>
- Fleury, D. (2021). *Questions en comment de raison : La révision des attentes du locuteur* [PhD]. Université de Paris/Université Paris Cité, Paris. https://u-paris.fr/theses/detail-dune-these/?id_these=3458
- Fleury, D. et Tovenà, L. (2018a). À propos des lectures des questions en comment. *SHS Web of Conferences*, 46, 12003. <https://doi.org/10.1051/shsconf/20184612003>
- Fleury, D. et Tovenà, L. (2018b). Reason questions with *comment* are expressions of an attributional search. Dans *Proceedings of the 22nd Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. SEMDIAL. http://semdial.org/anthology/Z18-Fleury_semdial_0015.pdf
- Fliessbach, J., Brunetti, L. et Yoo, H. (under review). On the overlapping discourse functions of Spanish 'cómo que' and French 'comment ça' interrogatives. Dans H. Metslang (dir.), *Open Linguistics. Subjectivity and Intersubjectivity in Language*. de Gruyter.
- Fontaine, J. J. R., Scherer, K. R. et Soriano, C. (dir.). (2013). *Series in affective science. Components of emotional meaning : A sourcebook*. Oxford University Press.
- Fournier, N. (1998). *Grammaire du français classique. Belin SUP Lettres*. Belin.
- Hamlaoui, F. (2011). On the role of phonology and discourse in Francilian French wh-questions. *Journal of Linguistics*, 47(1), 129–162. <https://doi.org/10.1017/S0022226710000198>
- Kédochim, F. et Guesser, S. (2020). On the readings of *pourquoi* in wh-questions. *Revista Lingvística*, 206–230.
- Kircher, R. (2016). The Matched-Guise Technique. Dans H. Zhu (dir.), *Guides to research methods in language and linguistics : Vol. 8. Research methods in intercultural communication : A practical guide* (p. 196–211). Wiley-Blackwell. <https://doi.org/10.1002/9781119166283.ch13>
- Kohler, K. J. (2005). Timing and communicative functions of pitch contours. *Phonetica*, 62(2-4), 88–105. <https://doi.org/10.1159/000090091>
- Korzen, H. (1990). Pourquoi *pourquoi* est-il différent ? L'adverbiale de cause et la classification des adverbiaux en général. *Langue Française*, 88(1), 60–79. <https://doi.org/10.3406/lfr.1990.5753>
- Lambert, W. E., Hodgson, R. C., Gardner, R. C. et Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, 60, 44–51. <https://doi.org/10.1037/h0044430>
- Lambrecht, K. (1981). *Topic, antitopic and verb agreement in non-standard French. Pragmatics & beyond : 2:6*. J. Benjamins.

- Liddell, T. M. et Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Massot, B. et Rowlett, P. (2013). Le débat sur la diglossie en France: Aspects scientifiques et politiques *Journal of French Language Studies*, 23(1), 1–16. <https://doi.org/10.1017/S0959269512000336>
- Myers, L. L. et Pellet, S. (2014). *Pourquoi* in Spoken French: Corpus-based function-form mapping. Dans S. Katz Bourns et L. L. Myers (dir.), *Pragmatics & Beyond New series : Vol. 244. Perspectives on Linguistic Structure and Context* (Vol. 244, p. 157–182). John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.244.08mye>
- Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, 49(3), 197–237. <https://doi.org/10.1037/h0055737>
- Posit team. (2023). *RStudio: Integrated Development Environment for R*. <http://www.posit.co/>
- Quillard, V. (2001). La diversité des formes interrogatives : comment l'interpréter ? *Langage Et Société*, 95(1), 57–72. <https://doi.org/10.3917/ls.095.0057>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>
- Smirnova, A. et Abeillé, A. (2021). Question particles ça and donc in French : A corpus study. *Linguistic Research*, 38(2), 239–269.
- Sorensen, T., Hohenstein, S. et Vasissth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, 12(3), 175–200. <https://doi.org/10.20982/tqmp.12.3.p175>
- Sullivan, G. M. et Artino, A. R. (2013). Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541–542. <https://doi.org/10.4300/JGME-5-4-18>
- Thiberge, G. (2020). *Acquisition et maîtrise des interrogatives partielles en français: La variation sociolinguistique comme outil interactionnel* [Doctoral Thesis]. Université de Paris, Paris.
- Tovena, L. M. (2023). Asking about the reason for an effect, and some consequences for the analysis of *wh*-interrogatives. Dans Ł. Jędrzejowski et C. Umbach (dir.), *Non-Interrogative Subordinate Wh-Clauses* (p. 381–409). Oxford University PressOxford. <https://doi.org/10.1093/oso/9780192844620.003.0012>
- Trotzke, A. et Czypionka, A. (2022). The pragmatics of surprise-disapproval questions: An empirical study. *Linguistics Vanguard*, 8(s2), 239–249. <https://doi.org/10.1515/lingvan-2020-0107>
- Tsai, W.-T. D. (2008). Left periphery and how-why alternations. *Journal of East Asian Linguistics*, 17, 83–115. <https://doi.org/10.1007/s10831-008-9021-0>

¹ Le mode subjonctif est optionnel dans cette phrase, comme c'est souvent le cas en français (Poplack *et al.* 2013). Il se peut que le choix du mode ait pourtant aussi des effets interprétatifs. Nous faisons l'hypothèse que l'indicatif favorise une lecture expressive de surprise et/ou reproche, alors que le subjonctif favoriserait une lecture de discours rapporté (« il paraît que ») qui motiverait une fonction de clarification de la question (donc demande d'information) plutôt qu'une fonction expressive.

² Le terme *pseudo-préjacente* ou *préjacente* par brévitité est utilisé par Fleury et Tovena (2018b) (et ensuite par Brunetti *et al.*, 2021; Brunetti *et al.*, 2022) dans leur analyse des interrogatives introduites par *comment* de raison. Fleury et Tovena empruntent le terme à la littérature sur la modalité et désignent avec celui-ci la proposition sémantique exprimée par la phrase sur laquelle s'applique l'élément interrogatif de raison (« the proposition expressed by the clause the *comment* of reason operates on », Fleury et Tovena, 2018b, p. 2). Contrairement aux interrogatives qui questionnent un dépendant du verbe, la préjacente des interrogatives à fonction de raison ne contient pas de variable. Fleury et Tovena (2018b) ajoutent le préfixe *pseudo-* parce qu'ils ne suivent pas une analyse où l'élément *wh-* serait un opérateur de phrase.

³ Dans Dekhissi et Coveney (2021, p. 134), cette classe de verbes regroupe les verbes marqués comme étant « familiers, populaires ou vulgaires » et les verbes en verlan.

⁴ L'utilisation de *qu'est-ce que* non-argumental peut être rapprochée de l'utilisation de *que* à fonction de circonstant, signifiant « pourquoi » en français classique, comme dans « Que tardez-vous, Seigneur, à la répudier ? » (Racine) (voir Fournier, 1998, p. 181). Nous remercions la ou le reviewer anonyme pour nous l'avoir indiqué.

⁵ La traduction est la nôtre.

⁶ Dans le cadre de cet article nous rapporterons les résultats statistiques principaux en précisant la valeur moyenne du coefficient estimé (Coef.Est., soit la valeur numérique caractérisant le lien entre la variable dépendante et la ou les variables indépendantes), la probabilité que ce coefficient soit inférieur à 0 ($P(\text{Coef.} < 0)$) ou supérieur à 0 ($P(\text{Coef.} > 0)$) afin d'évaluer la robustesse de cette évaluation, et un intervalle de crédibilité de 95% (95%CrI) qui indique les valeurs entre lesquelles il y a 95% de chances de trouver la valeur véritable du coefficient d'après les données récoltées.

⁷ SE = Erreur Type autour de la moyenne.