

Biagio Ursi, Carole Etienne
ICAR (CNRS UMR 5191) & ENS de Lyon

Florence Oloff
Faculté de sciences humaines –
Université d’Oulu (Finlande)

Lorenza Mondada
Séminaire d’études françaises –
Université de Bâle (Suisse)

Véronique Traverso
Institut français du Proche-Orient –
Beyrouth (Liban) & ICAR (CNRS UMR
5191)

Diversité des répétitions et des reformulations dans les interactions orales : défis analytiques et conception d’un outil de détection automatique

1. REFORMULATION ET RÉPÉTITION DANS L’INTERACTION

La reformulation a attiré l’attention des chercheurs dans de nombreuses perspectives en linguistique. Dans les études sur la reformulation dans la conversation, elle a souvent été identifiée grâce aux marqueurs qui introduisent ce qui est reformulé, tout en précisant le positionnement énonciatif des locuteurs par rapport à ce qui est dit. Ainsi, par exemple, les études menées dans le cadre de l’analyse du discours genevoise (Roulet 1987), ainsi que les travaux de C. Rossari (1989, 1992, 1994), distinguent les reformulations qui apportent une rectification ou clarification à ce qui a précédemment été dit par le locuteur (reformulations paraphrastiques) et les reformulations qui marquent une prise de distance par rapport à ce que le locuteur vient d’énoncer (reformulation non paraphrastiques). D’autres chercheurs ont proposé des analyses syntaxiques en s’appuyant également sur des marqueurs souvent, mais pas exclusivement, considérés comme des indices de reformulation et dont les caractéristiques sémantiques et pragmatiques spécifiques ont été discutées (v., entre autres, Le Bot, Schuwer & Richard 2008).

Dans la perspective interactionnelle, la prise en compte de la séquentialité des échanges constitue l’un des critères de classification les plus utilisés. D’après E. Gülich et T. Kotschi (1987 : 48-52), par exemple, quatre types de configurations sont possibles, en distinguant le locuteur qui accomplit la reformulation de celui qui déclenche le processus de reformulation. Ce dernier manifeste souvent un souci de clarification ou, dans la perspective de l’analyse conversationnelle, peut exploiter des pratiques de réparation (Jefferson 1972 ; Schegloff, Jefferson &

Sacks 1977 ; Couper-Kuhlen 1996 ; Curl 2005). Ainsi, dans la lignée de ces travaux, on distingue l'auto-reformulation (le locuteur A reformule ses propres propos) de l'hétéro-reformulation (le locuteur B reformule les propos du locuteur A) ; en outre, on distingue les reformulations auto-initiées des reformulations hétéro-initiées, selon le locuteur qui en prend l'initiative et la déclenche. Un statut préférentiel est généralement accordé aux auto-reformulations, « utilisées nettement plus souvent que les hétéro-reformulations » (Gülich & Kotschi, 1987 : 52). En particulier, les auto-reformulations auto-initiées représentent la majorité des cas, y compris dans les conversations entre locuteurs francophones non natifs (87 % des cas d'après l'étude de Kanaan, 2011 : 115).

Le phénomène de répétition, quant à lui, a été largement étudié dans le domaine des études interactionnelles sur le plan des fonctions, dont certaines s'apparentent aux reformulations (cf. l'inventaire proposé par Norrick 1987). Et les mêmes caractérisations séquentielles des reformulations peuvent être appliquées aux répétitions : on distingue ainsi l'auto-répétition (le locuteur A répète ce qu'il a dit précédemment) de l'hétéro-répétition (le locuteur B reproduit ce qui a été prononcé par le locuteur A).

Dans cet article, nous nous concentrons sur des formes particulières de reformulation : les répétitions caractérisées par une reprise plus ou moins *verbatim* d'un élément précédemment énoncé. Sur le plan formel, et toujours dans le domaine de l'interaction, la répétition et la reformulation sont étroitement liées. Ainsi, D. Tannen (1989 : 54) parle de *continuum* entre répétition *verbatim* (*exact repetition*) et reformulation proprement dite (*paraphrase*) : dans ce dernier cas, la répétition concerne uniquement le contenu de ce qui est dit, mais pas sa forme. Entre ces deux extrêmes, des degrés de variabilité sont observés : les formes de répétitions partielles peuvent se distinguer sur le plan lexical, morphologique et/ou syntaxique. Alors que la reformulation est traditionnellement comprise comme étant explicitement signalée par des connecteurs, la répétition, du fait de sa fréquence élevée dans le discours et de l'absence possible de tels marqueurs, représente une pratique de reformulation plus opaque et possédant des fonctionnalités différentes. Cette étude porte sur les reformulations que nous pouvons identifier par la répétition à partir d'un outil de détection automatique.

Afin d'illustrer les spécificités de cet outil, nous présenterons d'abord la base de données CLAPI, pour et à partir de laquelle l'outil a été conçu. Ensuite, nous nous pencherons sur la méthodologie de conception de l'outil automatique pour le repérage des hétéro-répétitions. Enfin, nous proposerons quelques pistes pour étudier l'hétéro-répétition comme ressource dans les pratiques de reformulation, en mobilisant cet outil.

Notre étude se fonde exclusivement sur des corpus de français parlé en interaction, et plus particulièrement des enregistrements de situations naturelles ¹, ce qui représente une spécificité par rapport aux autres contributions de ce volume. Nos corpus sont disponibles et requêtables en ligne, comme ceux étudiés par V. Magri (2018), C. Rossari, C. Ricci et D. Wandel (2018) pour l'écrit et par A. Steuckardt (2018) pour l'écrit et pour l'oral (tous trois dans ce volume). Cependant, contrairement aux travaux susmentionnés, notre contribution n'envisage pas la reformulation à partir de marqueurs spécifiques ou de formes prédéfinies. Nous proposons une recherche ciblée à partir d'attestations détectées automatiquement d'un phénomène dans la variété de ses occurrences, ce qui nous conduit à repérer des patterns et des fonctions émergentes dans la séquentialité des tours de parole. Nos analyses sont essentiellement qualitatives, mais l'outil permet aussi des relevés quantitatifs.

2. LES CORPUS ORAUX ET LA PLATEFORME « CLAPI »

L'analyse des interactions sociales a été enrichie au cours des dix dernières années par le développement de bases de données qui ont permis aux chercheurs d'accéder à un nombre de corpus de plus en plus important et diversifié (v. Bruxelles *et alii* (éds) 2009 ; Avanzi, Béguelin & Diemoz (éds) 2016). La base de données CLAPI (Corpus de LAngue Parlée en Interaction) ², développée au laboratoire ICAR, est dédiée au français parlé en interaction (Groupe ICOR 2008, 2009, 2016). Elle archive des données collectées dans des situations quotidiennes d'échange d'une grande diversité : en milieu professionnel, dans des commerces, en milieu médical, en contexte institutionnel ou bien dans un cadre privé. Il s'agit de données écologiques, *i.e.* enregistrées ou filmées dans le contexte où elles se produisent naturellement (Groupe ICOR 2016).

Les données collectées depuis une quarantaine d'années ont été réunies, décrites et transcrites dans la base CLAPI pour mettre à disposition de la communauté scientifique des corpus d'interactions authentiques, fondamentales pour l'étude du français parlé en interaction. La base, opérationnelle depuis 2005, comprend à ce jour 60 corpus, ce qui correspond à 300 enregistrements dûment décrits dont 140 sont requêtables – soit 63 h de données (dont 48 h téléchargeables).

Pour garantir l'interopérabilité de CLAPI, un export en Dublin Core/Olac des descripteurs et un export en TEI des descripteurs et du contenu de la transcription ont été réalisés. Une partie des données de CLAPI sont aussi disponibles sur la plateforme d'archivage ORTOLANG.

1. Pour notre étude, nous n'avons pas sélectionné des enregistrements d'interactions construites par le chercheur comme, par exemple, les entretiens ou les conversations sur un thème imposé.

2. Les extraits présentés dans cet article ont été transcrits selon la convention ICOR (icar.univ-lyon2.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf).

La plateforme CLAPI comprend également des outils de requête permettant à l'utilisateur d'interroger les transcriptions, depuis la simple recherche d'une forme jusqu'à l'élaboration d'une requête complexe, à partir d'une trentaine de critères combinant les métadonnées, le lexique et certains phénomènes de l'oral comme les chevauchements, les pauses ou la position dans le tour de parole. Notre recherche sur la reformulation s'appuie sur l'un de ces outils que nous présentons *infra*.

3. MÉTHODOLOGIE DE CONCEPTION D'UN OUTIL AUTOMATIQUE ADAPTÉ À L'ÉTUDE DE L'INTERACTION

3.1. Définition de l'objet de recherche

La conception d'un outil automatique repose sur la définition et la modélisation du phénomène que l'on souhaite détecter sachant qu'à l'oral, et plus particulièrement en interaction, un même phénomène peut être réalisé de plusieurs manières, dans une séquence plus ou moins longue, par un seul locuteur ou en collaboration entre plusieurs participants.

Les études interactionnelles s'inscrivent dans une approche qualitative où le chercheur constitue manuellement ou semi-automatiquement une collection d'attestations correspondant au phénomène recherché, pour se livrer ensuite à une analyse fine et détaillée des cas relevés. L'expert a ainsi l'habitude de sélectionner « manuellement » les occurrences du phénomène dans une ou plusieurs interactions grâce à un ensemble complexe de propriétés, à la fois structurelles, séquentielles et actionnelles, qui ne peuvent se réduire à des règles formelles qu'un ordinateur pourrait prédéfinir et traiter.

Les outils automatiques visent ainsi à aider le chercheur à constituer des collections et à lui permettre de travailler sur des jeux de données plus importants, davantage qu'à générer automatiquement des résultats définitifs. Dans la conception des outils, et notamment dans la définition des critères permettant de discriminer le phénomène visé parmi d'autres, l'approche théorique et analytique de l'expert est fondamentale. D'un point de vue méthodologique, on procède le plus souvent par itération pour définir les critères qui permettent d'encadrer au mieux l'objet recherché sans écarter des résultats proches, mais sans non plus générer trop de « bruit » (cf. § 3.4.2, ii), ce qui rendrait l'outil moins pertinent, voire inapplicable.

3.2. Les outils automatiques et l'étude de l'interaction

L'analyse des interactions se fonde sur l'analyse séquentielle des échanges qui se construisent en temps réel au fil des contributions, verbales et non verbales, des locuteurs. Par conséquent, les outils de requête à l'oral s'inscrivent dans la même dynamique : dans une approche interactionnelle, on ne traite pas les tours de parole comme s'ils constituaient un seul texte continu dans lequel on pourrait

chercher et comptabiliser les occurrences du phénomène étudié. Au contraire, l'outil est conçu pour respecter la séquentialité et l'alternance des locuteurs, en distinguant si le phénomène étudié concerne un ou plusieurs locuteurs, voire s'il est produit par plusieurs d'entre eux (Groupe ICOR 2018).

Par ailleurs, l'étude des interactions amène souvent à s'interroger sur la fréquence du phénomène étudié selon le type de situation (réunion de travail, repas en famille, visite guidée, etc.), son caractère formel ou informel, le nombre de locuteurs et le mode de communication (interactions en présence, au téléphone, médiées par écran).

Les outils de la plateforme CLAPI permettent de bénéficier de la variété des situations présentes dans la base et d'étudier les résultats obtenus dans ces différentes perspectives. Par contre, ces outils automatiques requièrent des critères particulièrement robustes pour couvrir l'ensemble de ces données et s'adapter à leur hétérogénéité.

3.3. Le paramétrage de l'outil

Pour éviter l'effet « boîte noire », où un outil produit un ensemble de résultats de manière plus ou moins opaque sans permettre à l'utilisateur d'en modifier le cours, nous avons laissé à l'utilisateur la possibilité de choisir le jeu de paramètres qui lui convient, en explicitant chaque paramètre par des exemples et en proposant une valeur par défaut sur la base de nos différentes expérimentations.

3.4. L'outil de détection des hétéro-répétitions

Pour notre étude de la reformulation, nous partirons des hétéro-répétitions, définies comme des répétitions de segments plus ou moins longs par des locuteurs successifs, identifiées à l'aide d'un outil dédié à leur repérage.

Par convention, nous parlerons de *tokens*, et non de mots ; le *token* étant la représentation du mot tel qu'il est produit par le locuteur et qui ne correspond pas forcément à une forme orthographique. Ainsi, les mots tronqués ou élidés (« 'fin » au lieu de « enfin ») sont également considérés comme des *tokens*.

3.4.1. L'identification d'une répétition

Dans les interactions naturelles, le nombre de formes lexicales différentes représente seulement 5 à 15 % du nombre total d'occurrences. Ainsi, on relève 15 % de formes différentes dans la situation privée « Apéritif Pois » et 9.5 % dans la réunion de travail entre architectes « Mosaic ». Un grand nombre de formes lexicales (85 % ou plus) est donc utilisé à plusieurs reprises dans une même conversation. Par conséquent, nous pouvons supposer un nombre considérable de répétitions ; cependant, il faudrait pouvoir distinguer les « vraies » répétitions – perçues comme telles par les locuteurs – des simples réutilisations de lexèmes fréquents. Nous avons donc retenu comme premier critère de définition

d'une répétition la distance entre deux occurrences d'un même segment, que nous appellerons « empan ».

D'un point de vue cognitif, si les mécanismes de maintien de l'information verbale ont fait l'objet de différents modèles (v., entre autres, Baddeley 1986 ; Cowan 1992), il nous est difficile de considérer une durée moyenne pendant laquelle un ensemble de *tokens* serait perçu comme répété. En effet, certaines interactions, où les locuteurs sont engagés dans des activités (comme, par exemple, un traitement médical, les déplacements d'un groupe dans une visite, la manipulation d'artefacts), comprennent des pauses importantes qui modifient la valeur de cette mesure, la rendant instable et non applicable à des situations différentes, voire au sein de la même interaction.

L'empan pourrait être délimité par le nombre de tours de parole, mais la taille des tours de parole varie en fonction du type d'interaction, du transcripateur ou du logiciel de transcription utilisé. Cette unité n'est donc pas assez robuste.

Nous lui avons préféré un autre critère, basé sur le nombre de *tokens* entre deux segments répétés. Compte tenu de la fréquence des répétitions à l'oral, nous sommes amenés à distinguer trois niveaux d'empan suivant la longueur du segment : un *token* seul (5 par défaut), un segment de moins de 3 *tokens* (100 par défaut) et un segment de 3 *tokens* et plus (200 par défaut).

Pour pallier le problème posé par les répétitions des mots les plus fréquents, nous proposons en option de les retirer des segments courts (moins de 3 *tokens*) en proposant plusieurs listes séparées – à l'exception de certains formats, particulièrement intéressants à l'oral, comme les marqueurs ou les répétitions d'une structure telle que « c'est le » ou « c'est pour ». Les listes des formes à exclure, proposées en option, sont les suivantes :

- les formes : *.h, inaud., inaudible, x, xx* ;
- les formes : *oui, non, si, ouais, mouais, mum, hm, bon* ;
- la forme *euh* ;
- les formes : *mais, ben, donc, hein, voilà, et, mais, oh, ah* ;
- les formes : *un, une, la, le, les, l', l', ces, ça, ces, cet, cette, des, de, du, d', dans, que, qui, qu', à, au, aux, où, ou, je, tu, il, nous, vous, ils, elles, on, etc.*

De plus, une autre option, basée sur le calcul des *Term Frequency-Inverse Document Frequency* (TF-IDF), permet de retirer des répétitions les mots discriminants pour une transcription donnée comme, par exemple, le lexique spécialisé typique d'une interaction professionnelle (c'est le cas du corpus « Réunion de conception en architecture – Mosaic » dans notre étude).

3.4.2. L'identification d'un segment

(i) Ajout/suppression de mots dans le segment

Ce type de variation, très fréquent à l'oral, introduit dans le segment répété³ des mots qui modifient de manière plus ou moins sensible les mots du segment d'origine (1) ou en suppriment certains (2) :

- (1) Corpus : Conversation familières – Visites ~ Clodif Q4
P **et tu t'es pas fait inscrire à Lyon**=
S =**et j' me suis pas fait réinscrire à Lyon**
- (2) Corpus : Bielefeld – situations de contact ~ Chez le directeur du foyer
DOR parce que j' voulais **rendre** la clé de: de **la chambre** neuf/
au premier euh: (.) étage à gauche \ (0.7)
c'est le couloir des femmes \ là à droite/ xx
(1.1)
DIR pourquoi/ **rendre la chambre**/

Par conséquent, nous avons introduit un paramètre fixant le nombre d'« intrus » tolérés dans un segment à 3 par défaut.

(ii) Ordre et positionnement des segments

Un autre type de variation est la modification de l'ordre des *tokens* dans le segment répété, comme dans l'exemple (3) :

- (3) Corpus : Conversations familières – Visites ~ Clodif Q4
S **vous viendrez manger la prochaine fois**
C et vous r'montez après à V. après
M et oui (.) bien sûr
P **la prochaine fois on viendra manger**

Nous avons proposé par défaut cette option en prenant le risque d'ajouter des attestations qui ne sont en fait pas des répétitions, c'est ce que l'on appelle « bruit » dans un outil automatique.

(iii) Orthographe adaptée

Les locuteurs produisent souvent le même segment en prononçant les *tokens* différemment, avec ou sans élision (4).

- (4) Corpus : Conversation familières – Visites ~ Clodif Q4
P **un petit** fond
P (inaud.)
C **un p'tit** coup

(iv) Lemmatisation

La lemmatisation permet de rapprocher des formes conjuguées (p. ex. dans le passage de *on* à *nous* – comme *on va* vs *nous allons* – ou, dans les dialogues, de

3. Les segments répétés et d'origine apparaissent en rouge dans l'affichage de la transcription de la base CLAPI. Dans les extraits de cet article, ils sont en gras.

je vais à tu vas / vous allez) ou de prendre en compte des changements de temps (p. ex. entre le futur simple et le futur proche, cf. (5)).

- (5) Corpus : Bielefeld – situations de contact ~ Chez le directeur du foyer
- | | |
|-----|---|
| H1 | bon mais comme il sort pas euh je repasserai
j' vous laisserai tout |
| DIR | ou alors vous laissez un p'tit un p'tit mot
n'importe quoi hein |
| H1 | remarquez je je vais quand même regarder |
| DI | ben allez voir hein alors j' vous dis |

Un paramètre « inclure la recherche des lemmes » est présent dans l'outil.

3.4.3. Les résultats affichés

Les résultats sont

- affichés en rouge dans la transcription, pour les rendre visibles dans leur séquentialité ;
- regroupés en segments selon leur fréquence ;
- regroupés par locuteurs, puis par fréquence.

Le nombre de répétitions par locuteur pourra être rapproché du nombre de *tokens* produits par ce même locuteur. Ainsi, on pourra disposer d'éléments de comparaison et évaluer si un locuteur fait souvent appel à la répétition en tant que ressource conversationnelle.

4. ANALYSES QUALITATIVES

Nous avons testé l'outil de détection des hétéro-répétitions, avec le paramétrage spécifié *supra* (§ 3.4), sur les transcriptions de quatre enregistrements documentant quatre situations différentes :

- Boulangerie rurale, corpus : Interactions commerciales – petits commerces ⁴ ;
- Chez le directeur du foyer, corpus : Bielefeld – situations de contact ;
- Mosaic, corpus : Réunion de conception en architecture ;
- Pois, corpus : Apéritif entre ami(e)s.

Dans cette section, nous proposons une caractérisation interactionnelle de certains formats d'hétéro-répétitions qui ont été repérés à l'aide de notre outil automatique.

4.1. L'hétéro-répétition pour exprimer son accord

La première configuration retenue est une reformulation qui module une évaluation (Pomerantz 1984) par le biais d'une modification du degré adjectival, dans ce cas une intensification (l. 1-8) :

4. Les noms des transcriptions sont ceux relevés dans la base de données CLAPI dans la liste des corpus. Pour « Boulangerie rurale », la transcription se compose de deux parties.

- (6) Corpus : Réunion de conception en architecture ~ Mosaic
- | | | |
|----|---|---|
| 01 | M | parce que c'est c'est petit là/ |
| 02 | C | alors après non mais c' qui c- |
| 03 | M | j' me rappelle le le passage là est |
| 04 | C | oui il est petit |
| 05 | M | et et |
| 06 | C | il est ridicule/= |
| 07 | M | =oui |
| 08 | C | et de toute façon c'est trop petit pour être un bar/ ça\ |

Dans cette réunion, des architectes discutent d'un détail sur le plan ; la référence est déictiquement marquée par un geste de pointage vers le même endroit (l. 1 et l. 8). Une des spécificités des hétéro-répétitions de ce corpus est observable à la ligne 8 : la reprise d'un élément qui sert de support pour une expansion. Ici, la dimension de l'espace indiqué, sa petitesse, est associée à une utilisation improbable (*c'est trop petit pour être un bar ça*, le geste de pointage est maintenu jusqu'à la fin du tour, marquée par l'élément disloqué *ça*).

Cette pratique conversationnelle est employée pour marquer l'accord et peut éventuellement accompagner une hétéro-répétition à l'identique qui a la même fonction (v. Traverso 2012), comme dans l'extrait suivant :

- (7) Corpus : Apéritif entre ami(e)s ~ Pois
- | | | |
|----|-----|---|
| 01 | JUL | est-ce que c'est simple à faire/ pa'ce que bon j'ai |
| 02 | | aucune idée comment on fait la crème |
| 03 | ROM | s- c'est |
| 04 | ANN | c'est pas [dur] |
| 05 | ROM | [nan] c'est pas dur hein |
| 06 | | [c'est long \] |
| 07 | ANN | [sauf que c'est] super long |

Dans cet extrait, Julie demande à ses invités, Romain et Anne, des informations concernant la préparation de la crème de marrons. La réponse à la question initiale (*est-ce que c'est simple à faire/*, l. 1) est produite d'abord par Anne (l. 4) et ensuite par Romain, qui répète la réponse de sa copine (l. 5). Il garde la parole et produit une autre évaluation, toujours introduite par *c'est*, qui présente un autre adjectif *long* (l. 6). Anne reformule cette dernière intervention par une intensification (*c'est super long*, l. 7), partiellement en chevauchement.

La reformulation d'un élément, notamment à l'aide d'un modificateur, peut également apparaître dans l'hétéro-répétition exprimant l'accord. Cette fois l'« intrus » n'est pas inséré mais se trouve à une extrémité du segment répété – par exemple, à l'initiale :

- (8) Corpus : Réunion de conception en architecture ~ Mosaic
- | | | |
|----|---|---|
| 01 | C | et dans cette partie-là de sous-sol on avait récupéré |
| 02 | | cet espace-là pour faire des p'tites chambres de moines= |
| 03 | M | =ah oui e- |
| 04 | | (0.2) |
| 05 | C | .H donc euh qu' étaient euh: |

06	L	très [très p'tites]
07	C	[.h euh:][:]
08	M	[ouais]: très en longueur
09		[comme on les avait]&
10	C	[tout en longueur oui]
11	M	&faits [oui par] là
12	C	[v:oilà]

Nous retrouvons ici l'hétéro-répétition de l'adjectif, qui est intensifié pour exprimer l'accord avec un tour précédent (l. 2 et l. 6). La description de l'espace récupéré par les architectes et réaménagé pour un nouvel emploi est formulée à l'aide du syntagme *en longueur*, d'abord précédé par *très* (l. 8) et ensuite par *tout* (l. 10). Ici, la co-construction des commentaires sur le travail de conception est interactionnellement réalisée par les participants : le *oui* qui suit l'hétéro-répétition et le *voilà* chevauchant final sont des signaux conversationnels du travail collaboratif des architectes (v. Bruxelles, Greco & Mondada 2009 ; Traverso 2009).

4.2. L'hétéro-répétition pour effectuer une réparation

La « réparation » (*repair* en anglais) est l'un des phénomènes les plus courants à l'oral. En analyse conversationnelle, on range sous cette notion toute procédure visant la résolution de problèmes de compréhension qui émergent dans l'interaction. Contrairement à ce que l'on pourrait penser, la réparation ne se limite pas aux corrections d'erreurs de grammaire. La réparation touche aux hésitations, aux recherches de mots, mais aussi à tout type de problème intervenant dans un contexte spécifique⁵. L'hétéro-répétition peut servir d'initiation de la réparation, pour signaler un problème dans un tour précédent (Jefferson 1987), mais peut aussi représenter la réparation même d'un segment antérieur.

Dans l'extrait (9) *infra*, la vendeuse effectue une réparation suite au tour précédent de son fils, Marius (l. 3-4). En particulier, elle exploite la même construction verbale (*j' marque / tu marques*) et propose une formulation succincte pour les produits (*gel javel* (0.7) *un lit' ajax* (inaud.) / *juste ajax et gel javel c'est tout*) à noter sur papier. Dans la reformulation, l'objet de la construction verbale, qui représente le focus de l'activité d'écriture de l'enfant, est entouré par des marqueurs qui soulignent son caractère succinct (*juste... c'est tout*) :

- (9) Corpus : Interactions commerciales – petits commerces ~ Boulangerie rurale
- | | | |
|----|-----|--|
| 01 | MAR | regarde mama:n (0.8) un lit' / j' marque / (.) gel javel |
| 02 | | (0.7) un lit' ajax [(inaud.)] |
| 03 | VE1 | [oui 'fin] tu marques juste ajax |
| 04 | | et gel javel c'est tout \ ajax euh multi-usage et puis |
| 05 | | ((VE1 retourne au comptoir)) |
| 06 | MAR | et j' croyais qu' tu m'avais dit marquer un lit' |

5. V. Kitzinger (2013) pour l'anglais et Traverso (2016 : 101-119) pour le français.

- (12) Corpus : Apéritif entre ami(e)s ~ Pois
- | | | |
|----|-----|---|
| 01 | JUL | tu veux quoi en fait tu veux du jus d'orange/ jus: euh (.) |
| 02 | | multivitaminé / coca/ orangina/ j'ai aussi de l'eau |
| 03 | ANN | moi j' veux bien multivitaminé \ |
| 04 | JUL | <((rapide et à voix basse)) multivitaminé > |
| 05 | ANN | [ouais] |
| 06 | JUL | [d'ailleurs] romain tu veux un truc alcoolisé/ |
| 07 | | (0.6) |
| 08 | ROM | [euh::] |
| 09 | JUL | [xx encore] j'ai d' la vodka |
| 10 | | (0.5) |
| 11 | JUL | crè::me de [cassis/]& |
| 12 | ROM | [((rires))] |
| 13 | JUL | & ben non [fin avant faut aller à (inaud.) 'fin] |
| 14 | ROM | [non j' vais prendre euh: du] |
| 15 | | jus multivitaminé |
| 16 | JUL | multivitaminé bon |

La formulation succincte peut également servir d'accusé de réception, comme dans l'exemple (13), suite à une demande du client dans la boulangerie (l. 3) :

- (13) Corpus : Interactions commerciales – petits commerces ~ Boulangerie rurale
- | | | |
|----|-----|--|
| 01 | C05 | euh: une baguette/ (.) courte s'il vous plaît/ (.) |
| 02 | | °et un pain au chocolat° |
| 03 | VE1 | un pain choc |

Un autre emploi de la répétition qui manifeste la réception d'un élément d'un tour précédent est la complétion syntaxique d'un tour de parole par un autre participant : le locuteur qui a produit le tour incomplet peut accuser réception de la complétion et, en poursuivant son tour, la valider (*i.e.* la considérer comme adéquate, v. Oloff 2014). En voici un exemple :

- (14) Corpus : Réunion de conception en architecture ~ Mosaic
- | | | |
|----|---|--|
| 01 | C | .h euh ça c'était les p'tites chambres de:: de:: de[::]: |
| 02 | L | [moines] |
| 03 | | (0.2) |
| 04 | C | de moines h qu' étaient donc/ euh\ finalement comme une |
| 05 | | espèce de cabine de chemin de fer hein/ |

Dans ce cas, il s'agit plutôt d'une pratique de co-construction, liée à l'hétéro-répétition comme ressource pour la (résolution de la) formulation d'un tour de parole, et non pas pour une reformulation proprement dite.

Par ailleurs, toujours dans le corpus de réunion, le travail de co-construction peut être lié à l'hétéro-répétition, à considérer comme une formulation alternative plutôt que comme une reformulation après coup. C'est notamment le cas des productions chorales qui, contrairement à la configuration de l'exemple (14), sont réalisées conjointement par plusieurs participants (Lerner 2002) :

- (15) Corpus : Réunion de conception en architecture ~ Mosaic
- 01 C voilà [ce qu'on v]oit là qui permettait .h de **donner**/
 02 L [hm hm]
 03 (0.6)
 04 C mais c'était l'idée d'avoir quand même
 05 un a[pport de **lumière** nat-]
 06 L [**donner** d' la **lumière**:]
 07 C &d'un na- naturelle
 08 M d'accord

Ligne 1, le tour de C est incomplet. Le locuteur redémarre avec une construction différente qui exprime le concept envisagé au début de l'extrait (*mais c'était l'idée d'avoir quand même un apport de lumière nat-*, l. 4-5). Ce nouveau tour est chevauché par L, qui produit une complétion du tour initial (*donner d' la lumière*, l. 6) et contribue ainsi à la discussion de la partie du plan en question. Dans ce cas, l'outil permet de repérer deux segments impliqués dans la reformulation, où le deuxième s'inscrit dans une trajectoire initialement tracée par l'un des participants.

5. ANALYSES QUANTITATIVES

Dans les corpus sélectionnés, nous avons considéré des types d'interactions différents et présenté des configurations séquentielles qui répondent à des finalités spécifiques dans l'interaction. Dans cette section, nous nous appuyerons sur les tableaux des résultats affichés dans le CLAPI, qui offrent un aperçu quantitatif de la totalité des *tokens* hétéro-répétés et des locuteurs qui les produisent. Nous nous limiterons à trois observations sur trois enregistrements, qui serviront de pistes pour des considérations plus générales.

Token en répétition	Nombre total 44	LOCUTEUR DIR 30 soit 68%	LOCUTEUR DOR 13 soit 30%	LOCUTEUR H1 1 soit 2%
Nombre de tokens dans la transcription	2116	1430 soit 68%	660 soit 31%	26 soit 1%

Figure 1 : Bilan des tokens hétéro-répétés en inter-tour
 [corpus : Bielefeld – situations de contact ~ Chez le directeur du foyer]

Le tableau de la Figure 1 fournit le nombre de *tokens* hétéro-répétés par rapport aux *tokens* produits par les deux participants qui parlent dans la presque totalité de cet enregistrement. D'abord, on observe que le nombre de répétitions est en adéquation avec le nombre de contributions des deux participants. Si l'on croise ces résultats avec les métadonnées du corpus, on s'aperçoit que Doris, l'étudiante allemande qui ramène la clé au directeur du foyer, est le participant qui effectue le moins d'hétéro-répétitions, et ce, malgré le fait qu'elle ne soit pas de langue maternelle française. Ainsi, Doris n'utilise pas la répétition comme une ressource récurrente dans les cas d'une mauvaise compréhension de la langue, comme on pourrait le croire chez les non-natifs.

Si l'on considère le même tableau pour l'enregistrement Mosaic (Fig. 2) ⁶, toujours à partir de notre paramétrage, on s'aperçoit que la répartition des *tokens* hétéro-répétés par locuteur sur la totalité des *tokens* produits par les participants est proportionnelle. Elle ne semble pas représenter une stratégie privilégiée d'un interactant en particulier, malgré les degrés d'implication différents des participants dans le projet (notamment, « Marie [= locuteur M dans le tableau] a participé au projet dans une phase antérieure mais n'est pas au courant des évolutions récentes », Détienne & Traverso, 2009 : 67).

Token en répétition	Nombre total 189	LOCUTEUR C 87 soit 46%	LOCUTEUR L 52 soit 28%	LOCUTEUR M 50 soit 26%
Nombre de tokens dans la transcription	17650	9182 soit 52%	4798 soit 27%	3670 soit 21%

Figure 2 : Bilan des *tokens* hétéro-répétés en inter-tour [corpus : Réunion de conception en architecture ~ Mosaic]

Dans l'interaction à la boulangerie ⁷, si l'on parcourt les premières attestations hétéro-répétées par fréquence décroissante, on remarque un pourcentage significatif de formes figées, telles que les salutations (*bonjour, bonne journée, (au) r`voir*), en ouverture et clôture d'interaction, ainsi que des formules de politesse (*merci, s'il vous plaît*, cf. Fig. 3). Ces derniers éléments accompagnent les requêtes de produits par les clients et le rendu de la monnaie par la vendeuse.

Token en répétition	Nombre total 136
Nombre de tokens dans la transcription	1793
bonjour	16
merci	14
s' il vous plaît	14
bonne journée	12
r`voir	9
s' plaît	7
et	5
baguette longue	4
longue	4

Figure 3 : Bilan par fréquence décroissante des *tokens* hétéro-répétés en inter-tour [corpus : Interactions commerciales ~ Boulangerie rurale]

6. Par ailleurs, les mêmes considérations pourraient être formulées pour l'enregistrement *Pois*.

7. Nous considérons les neuf premiers *tokens* hétéro-répétés dans la première partie de la transcription, à titre d'exemple.

Ces résultats nous amènent à trois considérations qui peuvent fournir des pistes de recherche pour la langue parlée en interaction.

Premièrement, l'hétéro-répétition, et en particulier l'hétéro-répétition comme stratégie de reformulation, n'est pas une ressource mobilisée de manière prioritaire par des interactants qui ont un bagage linguistique (lexical ou grammatical) limité par rapport aux interactants dont le français est la langue maternelle.

Deuxièmement, dans une situation naturelle (qu'il s'agisse d'une réunion professionnelle ou d'une conversation entre amis) entre locuteurs natifs, l'hétéro-répétition n'apparaît pas forcément comme une ressource exploitée de manière privilégiée par un participant.

Enfin, le type de situation – comme par exemple une interaction dans un petit commerce – a un impact considérable sur la fonction de certaines formes d'hétéro-répétitions, qui sont utilisées dans des positions séquentiellement attendues. C'est notamment le cas des formes rituelles : les salutations en ouverture et clôture d'échange, et des mots de politesses dans le corpus enregistré dans une boulangerie rurale.

6. CONCLUSION

Dans cette contribution, nous avons abordé la reformulation sous l'angle des d'hétéro-répétitions. Nous avons détaillé les éléments qu'une définition de ce phénomène doit prendre en compte, concernant non seulement la diversité de leurs formes (répétition partielle ou complète, modifiée ou non) mais aussi de leurs fonctionnalités (notamment l'accord, la réparation et la réception).

Ces caractérisations deviennent particulièrement cruciales lorsqu'il s'agit de concevoir un outil de recherche automatique des répétitions, puisqu'il convient de déterminer ce que l'on souhaiterait qu'il puisse repérer d'un point de vue interactionnel. Ainsi, les répétitions modifiées, *i.e.* tout type de variation dans la reprise d'une production verbale initialement produite, constituent un champ de recherche privilégié pour explorer les fonctions des reformulations qui présentent différents degrés de variabilité formelle. Nous n'avons pas pu présenter tout l'éventail des configurations interactionnelles pour lesquelles les hétéro-répétitions contribuent à définir des reformulations ⁸.

La possibilité de repérer des hétéro-répétitions contenant des « intrus » est particulièrement intéressante pour qualifier d'un point de vue formel ces reformulations, qui sont des reprises caractérisées par l'ajout de mots (à l'intérieur

8. Par exemple, on pourrait ajouter les séquences de question-réponse (JUL : « t' **as fait les réservations** pour l' resto nan/ ROM : euh nan\ j'**ai pas fait les réservations** ») ou les emplois de citation (comme dans un échange de l'interaction à la boulangerie, où la vendeuse sert d'intermédiaire entre les souhaits de la cliente C35 et la parole du boulanger qui est dans la pièce à côté). Dans ces cas, la recherche par lemme s'avère centrale.

du segment répété ou non). En revanche, les hétéro-répétitions que nous avons relevées *supra* (§ 4.3) constituent davantage des reprises partielles pour accuser réception d'un terme précédemment mentionné par le locuteur. Dans ce cas, elles ne sont pas assimilables aux reformulations introduites par des « connecteurs pragmatiques », au sens d'E. Gülich et T. Kotschi (1983 : 309), mais peuvent être interprétées comme des « reprises », sans une véritable implication énonciative des locuteurs qui reformulent (Vion 2006) ⁹.

La prise en compte du contexte situationnel mérite une attention particulière pour l'étude des répétitions. Les résultats quantitatifs obtenus avec l'outil automatique ici présenté (§ 3.4) pour les segments hétéro-répétés nous offrent quelques pistes de réflexion. L'étude outillée des fonctions des hétéro-répétitions devrait donc se traduire par un paramétrage qui s'ajuste aux spécificités du cadre de l'interaction, à l'aide d'un va-et-vient entre l'outil automatique et les analyses qualitatives. Une approche multidimensionnelle est donc indispensable (Eshkol-Taravella & Grabar 2016), non seulement en aval mais aussi en amont ¹⁰, pour améliorer les performances de l'outil ainsi que son ergonomie.

Références

- [CLAPI] *Corpus de Langue Parlée en Interaction*, ICAR (CNRS UMR 5191 & Université Lyon 2). [clapi.ish-lyon.cnrs.fr/]
- [ORTOLANG] *Outils et Ressources pour un Traitement Optimisé de la LANGue*. [www.ortolang.fr]
- AVANZI M., BÉGUELIN M.-J. & DIÉMOZ F. (éds) (2016), *Corpus n° 15 : Corpus de français parlés et français parlé des corpus*. [disponible en ligne]
- BADDELEY A. D. (1986), *Working Memory*, Oxford, Oxford University Press.
- BRUXELLES S. et alii (éds) ([2007] 2009), *Cahiers de linguistique n° 33 (2) : Grands corpus de français parlé : bilan historique et perspectives de recherche*, Louvain-la-Neuve, Éditions E.M.E.
- BRUXELLES S., GRECO L. & MONDADA L. (2009), « Pratiques de transition : ressources multimodales pour la structuration de l'activité », dans F. Détienne & V. Traverso (éds), *Méthodologies d'analyse de situations coopératives de conception : Corpus MOSAIC*, Nancy, Presses universitaires de Nancy, 221-301.
- CISLARU G. & OLIVE T. (2018), « Les jets textuels de révision : un point de vue dynamique sur la reformulation », *Langages* 212. (ce volume)

9. Pour Vion, « reprise » est le terme générique désignant la pure et simple répétition d'un segment textuel aussi bien que les différents degrés de ses reformulations (2006 : 11).

10. Les indices prosodiques soigneusement transcrits (courbes mélodiques, saillances, pauses) pourraient représenter des éléments utiles pour caractériser certains types de reformulations et contribueraient au repérage de patterns spécifiques. Dans les quatre enregistrements ici analysés, les pauses qui précèdent les hétéro-répétitions ne sont pas significatives. En revanche, dans les corpus d'écriture en temps réel, les jets textuels de reformulation sont précédés de pauses plus longues et ces jets sont produits légèrement plus lentement que ceux de production (Cislaru & Olive 2018 ici-même). Steuckardt (2018 ici-même) envisage la prise en compte de ces « indicateurs » à l'oral (et des marques typographiques à l'écrit) pour apprécier l'importance relative des différentes configurations auxquelles le segment reformulé, le segment reformulant et le marqueur peuvent donner lieu.

- COUPER-KUHLEN E. (1996), "The prosody of repetition: On quoting and mimicry", in E. Couper-Kuhlen & M. Selting (eds.), *Prosody in Conversation*, Cambridge, Cambridge University Press, 366-405.
- COWAN N. (1992), "Verbal memory span and the timing of spoken recall", *Journal of Memory and Language* 31 (5), 668-684.
- CURL T. S. (2005), "Practices in other-initiated repair resolution: The phonetic differentiation of 'repetitions'", *Discourse Processes* 39 (1), 1-43.
- DÉTIENNE F. & TRAVERSO V. (2009), « Présentation du corpus et de la situation », dans F. Détienne & V. Traverso (éds), *Méthodologies d'analyse de situations coopératives de conception : Corpus MOSAIC*, Nancy, Presses universitaires de Nancy, 61-68.
- ESHKOL-TARAVELLA I. & GRABAR N. (2016), « Reformulation à l'oral et dans le forum Web », dans F. Neveu et alii (éds), *5^e Congrès Mondial de Linguistique Française – CMLF 2016* (Tours, France), Les Ulis, EDP Sciences. [disponible en ligne]
- GROUPE ICOR (2008), "Tool-assisted analysis of interactional corpora: *voilà* in the CLAPI database", *Journal of French Language Studies* 18 (1), 121-145.
- GROUPE ICOR (2009), « Exploitation de la plateforme Corpus de Langue Parlée en Interaction (CLAPI) : le cas de *voilà* dans les chevauchements », *Cahiers de linguistique* 33 (2), 243-268.
- GROUPE ICOR (2016), « CLAPI, une base de données multimodale pour la parole en interaction : apports et dilemmes », *Corpus* 15, 165-191.
- GROUPE ICOR (2018), « Les hétéro-répétitions dans une interaction orale : définition et conception d'un outil de détection automatique », dans É. Richard (éd.), *Des organisations « dynamiques » de l'oral*, Berne, Peter Lang, 163-178.
- GÜLICH E. & KOTSCHI T. (1983), « Les marqueurs de la reformulation paraphrastique », *Cahiers de linguistique française* 5, 305-351.
- GÜLICH E. & KOTSCHI T. (1987), « Les actes de reformulation dans la consultation *La dame de Caluire* », dans P. Bange (éd.), *L'analyse des interactions verbales. La dame de Caluire : une consultation*, Berne, Peter Lang, 15-81.
- JEFFERSON G. (1972), "Side sequences", in D. Sudnow (ed.), *Studies in Social Interaction*, New York, The Free Press, 294-338.
- JEFFERSON G. (1987), "On exposed and embedded correction in conversation", in G. Button & J. R. E. Lee (eds.), *Talk and Social Organization*, Clevedon (UK), Multilingual Matters, 86-100.
- KANAAN L. (2011), *Reformulations, contacts de langues et compétence de communication : analyse linguistique et interactionnelle dans des discussions entre jeunes Libanais francophones*, Thèse de l'Université d'Orléans.
- KITZINGER C. (2013), "Repair", in J. Sidnell & T. Stivers (eds.), *The Handbook of Conversation Analysis*, Malden (MA), Wiley-Blackwell, 229-256.
- LE BOT M.-C., SCHUWER M. & RICHARD É. (éds) (2008), *La reformulation : marqueurs linguistiques, stratégies énonciatives*, Rennes, Presses universitaire de Rennes.
- LERNER G. H. (2002), "Turn-Sharing: The coral co-production of talk-in-interaction", in C. E. Ford, B. A. Fox & S. A. Thompson (eds.), *The Language of Turn and Sequence*, New York, Oxford University Press, 225-256.
- MAGRI V. (2018), « Marqueurs de reformulation : exploration outillée et contrastive dans deux corpus narratifs », *Langages* 212. (ce volume)
- NORRICK N. (1987), "Functions of repetition in conversation", *Text* 7 (3), 245-264.

- OLOFF F. (2014), « L'évaluation des complétions collaboratives : analyse séquentielle et multi-modale de tours de parole co-construits », dans F. Neveu *et alii* (éds), *4^e Congrès Mondial de Linguistique Française – CMLF 2014* (Berlin, Allemagne), Les Ulis, EDP Sciences, 2125-2145. [disponible en ligne]
- POMERANTZ A. (1984), "Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes", in J. M. Atkinson & J. Heritage (eds.), *Structures of Social Actions: Studies in Conversation Analysis*, Cambridge, Cambridge University Press, 57-101.
- ROSSARI C. (1989), « Des apports de l'analyse contrastive à la description de certains connecteurs reformulatifs du français et de l'italien », *Cahiers de linguistique française* 10, 193-214.
- ROSSARI C. (1992), « De l'exploitation de quelques connecteurs reformulatifs dans la gestion des articulations discursives », *Pratiques* 75, 111-124.
- ROSSARI C. (1994), *Les opérations de reformulation : analyse du processus et des marques dans une perspective contrastive français-italien*, Berne, Peter Lang.
- ROSSARI C., RICCI C. & WANDEL D. (2018), « Introduceurs de cadres et connecteurs de reformulation : étude contrastive sur corpus », *Langages* 212. (ce volume)
- ROULET E. (1987), « Complétude interactive et connecteurs reformulatifs », *Cahiers de linguistique française* 8, 111-140.
- SCHEGLOFF E. A., JEFFERSON G. & SACKS H. (1977), "The preference for self-correction in the organization of repair in conversation", *Language* 53 (2), 361-382.
- STEUCKARDT A. (2018), « Les marqueurs de reformulation formés sur *dire* : exploration outillée », *Langages* 212. (ce volume)
- TANNEN D. (1989), *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*, Cambridge, Cambridge University Press.
- TRAVERSO V. (2009), « Co-élaboration de solutions et rôle du graphico-gestuel : analyse interactionnelle », dans F. Détienne & V. Traverso (éds), *Méthodologies d'analyse de situations coopératives de conception : Corpus MOSAIC*, Nancy, Presses universitaires de Nancy, 87-182.
- TRAVERSO V. (2012), « Organisation du cadre participatif, accord et répétition dans l'interaction », dans F. Neveu *et alii* (éds), *3^e Congrès Mondial de Linguistique Française – CMLF 2012* (Lyon, France), Les Ulis, EDP Sciences, 663-679. [disponible en ligne]
- TRAVERSO V. (2016), *Décrire le français parlé en interaction*, Paris, Ophrys.
- VION R. (2006), « Reprise et mode d'implication énonciative », *La linguistique* 42, 11-28.