

Université de Montréal

La coordination dans différents contextes : comparaison entre la vidéoconférence et la réalité virtuelle

Par  
Émilie Lamarre

École d'orthophonie et d'audiologie, Faculté de médecine

Mémoire présenté en vue de l'obtention du grade de maîtrise ès sciences (M.Sc.)  
en sciences de l'orthophonie et de l'audiologie

Août 2023

© Émilie Lamarre, 2023

Université de Montréal

École d'orthophonie et d'audiologie, Faculté de médecine

---

*Ce mémoire intitulé*

**La coordination dans différents contextes : comparaison entre la vidéoconférence et la réalité virtuelle**

*Présenté par*

**Émilie Lamarre**

*A été évalué par un jury composé des personnes suivantes*

**Carole Anglade**

Président-rapporteur

**Stefano Rezzonico**

Directeur de recherche

**Douglas Shiller**

Codirecteur

**Simone Girard-Groeber**

Membre du jury

## Résumé

Sachant que les différents contextes d'interaction ont un impact sur la manière dont la conversation se déroule, le présent mémoire se concentre sur les moyens de communication à distance les plus récents : la vidéoconférence et la réalité virtuelle (VR). Nous posons la question suivante : en quoi la coordination de la conversation en réalité virtuelle est-elle différente de celle en vidéoconférence ?

Pour répondre à cette question, nous poursuivons deux objectifs : (1) comparer la fréquence et la durée des silences et des chevauchements dans les différents contextes et (2) décrire les moments où il y a un bris dans la fluidité de la conversation dans chaque contexte, puis comparer ces événements entre les contextes.

Nos résultats montrent que la VR n'influence pas la coordination de la conversation de la même manière que la vidéoconférence. Entre autres, les silences en VR sont plus longs et plus nombreux. De plus, les silences et les chevauchements en VR surviennent à des moments différents et ils se déroulent différemment. Les différences entre la VR et la vidéoconférence pourraient s'expliquer, entre autres, par la latence plus grande en VR, la plus grande familiarité des individus avec le contexte en vidéoconférence, les limites de la VR dans la transmission de l'information visuelle fine et le sentiment de coprésence qui se manifeste différemment dans chacun des contextes.

Notre étude apporte une compréhension nuancée des interactions en vidéoconférence et en VR et souhaite encourager le développement de la recherche sur ces contextes afin d'en promouvoir une utilisation plus avisée.

**Mots-clés** : Interaction sociale, Réalité virtuelle, Vidéoconférence, Pragmatique, Coordination, Tour de parole.

## Abstract

We know that conversations are influenced by the context in which they take place. Therefore, this paper focuses on how conversation is managed in new forms of remote communication: videoconference and virtual reality (VR). We ask this question: how is the coordination in VR different than the coordination in videoconference?

To answer this question, we pursue two objectives: (1) to compare frequency and duration of silences and overlaps between the contexts and (2) to describe what happens when the conversation flow is disturbed and then to compare those events between the contexts.

Our results show that VR influences coordination between interlocutors differently than videoconference. Silences in VR are longer and happen more often than in videoconference. Moreover, silences and overlaps take place at different points in VR conversations and they unfold in a different way. The differences between VR and videoconference might be explained by the fact that there is a greater latency in VR, that participants are less familiar with the context of VR, that VR is limited in its transmission of non-verbal communication and that VR has a different effect on social presence.

Our study sheds light on the importance of learning more about this new context of interaction (VR) before using it on a larger scale for interventions (e.g. with different clinical populations). Our conclusion can serve as a starting point for researchers who wish to contribute to a better understanding of the consequences of using VR as a means of communication.

**Keywords** : Social interaction, Virtual Reality, Videoconference, Pragmatics, Coordination, Turn-taking.

# Table des matières

Résumé .....	3
Abstract.....	4
Table des matières.....	5
Liste des tableaux .....	8
Liste des figures .....	9
Liste des sigles et abréviations.....	10
Remerciements.....	12
Chapitre 1 – Introduction.....	13
Chapitre 2 – Contexte théorique .....	16
2.1    L’analyse conversationnelle .....	16
2.2    Le tour de parole.....	17
2.2.1    Le système du tour de parole .....	17
2.2.2    Les facteurs qui influencent le tour de parole .....	20
2.3    La coordination entre les partenaires de communication .....	21
2.3.1    Les silences.....	22
2.3.2    Les chevauchements.....	24
2.4    Les contextes d’interaction .....	26
2.4.1    La vidéoconférence .....	26
2.4.2    La réalité virtuelle .....	29
Chapitre 3 – Objectifs et hypothèses.....	32
3.1    Premier objectif : analyse quantitative.....	32
3.2    Second objectif : analyse qualitative.....	33

Chapitre 4 – Méthodologie .....	34
4.1 Échantillonnage.....	34
4.2 Éthique .....	34
4.3 Procédure.....	35
4.4 Collecte des données .....	36
4.5 Méthodes d’analyses .....	37
4.5.1 Préparation des données .....	37
4.5.1.1 Les silences.....	38
4.5.1.2 Les chevauchements .....	38
4.5.2 Analyses quantitatives .....	38
4.5.3 Analyses qualitatives.....	41
4.5.3.1 Événements problématiques ciblés par les analyses qualitatives.....	43
Chapitre 5 – Résultats .....	45
5.1 Échantillon .....	45
5.2 Les contextes d’interaction et les indices de la coordination .....	46
5.2.1 Les silences.....	46
5.2.1.1 La fréquence des silences.....	46
5.2.1.2 La durée des silences .....	48
5.2.2 Les chevauchements .....	49
5.2.3 Conclusion sur le premier objectif .....	50
5.3 Les contextes d’interaction et les événements problématiques .....	51
5.3.1 Les silences.....	51
5.3.1.1 Causes apparentes des longs silences.....	52
5.3.1.2 L’interaction pendant les silences.....	56

5.3.2	Les chevauchements .....	60
5.3.2.1	Le contexte d'apparition du chevauchement .....	61
5.3.2.2	Le type de chevauchement .....	63
5.3.2.3	Le dénouement du chevauchement .....	66
5.3.3	Conclusion sur le second objectif.....	67
Chapitre 6 – Discussion .....		69
6.1	La latence .....	70
6.2	La familiarité .....	71
6.3	Les gestes et les expressions faciales.....	73
6.4	La coprésence .....	75
6.5	Conclusion de la discussion.....	77
Chapitre 7 – Conclusion .....		78
7.1	Les limites de la recherche et réflexions pour les recherches futures.....	78
7.2	Les retombées pour le domaine de la pratique .....	80
Références bibliographiques .....		82
Annexe A.....		92

## Liste des tableaux

<b>Tableau 1</b>	<i>Données descriptives des conversations en fonction des dyades.....</i>	45
<b>Tableau 2</b>	<i>Distribution et durée de tous les silences .....</i>	47
<b>Tableau 3</b>	<i>Distribution et durée des longs silences .....</i>	48
<b>Tableau 4</b>	<i>Distribution des chevauchements selon les contextes d'interaction .....</i>	50
<b>Tableau 5</b>	<i>Synthèse des thèmes et des catégories des longs silences.....</i>	52
<b>Tableau 6</b>	<i>Synthèse des thèmes et des catégories de chevauchements .....</i>	61
<b>Tableau 7</b>	<i>Listes des tâches de conversation par paires .....</i>	92



## Liste des figures

<b>Figure 1</b>	<i>Exemple d'un problème engendré dans la réalité virtuelle .....</i>	55
<b>Figure 2</b>	<i>Position réelle du participant P1 .....</i>	55
<b>Figure 3</b>	<i>Exemple de maintien de l'interaction pendant le long silence grâce aux gestes .....</i>	59
<b>Figure 4</b>	<i>Exemple de gestes qui ne maintiennent pas l'interaction durant le long silence .....</i>	59

## Liste des sigles et abréviations

ANOVA : Analyse de la variance

BC : Régulateurs (backchannel)

BF : Facteur de Bayes

M : Moyenne

N : Nombre d'occurrences

TDL : Trouble développemental du langage

TRP : Moment propice à la transition (transition-relevance place)

VR : Réalité virtuelle

*À Coralie, mon modèle d'ambition,  
de ténacité et de persévérance*

## Remerciements

J'aimerais remercier mes directeurs, Stefano Rezzonico et Douglas Shiller, pour leur soutien et leur confiance durant mes années à la maîtrise. Leurs lectures et leurs commentaires, toujours faits avec bienveillance, m'ont guidé dans mes réflexions. Non seulement ils m'ont permis de mener un projet de qualité, mais ils m'ont aussi aidé à me développer en tant que chercheuse.

Je remercie les membres du jury qui ont évalués mon mémoire et qui m'ont permis, par leurs commentaires judicieux, d'en améliorer la qualité.

Je remercie aussi Philippe, qui m'a soutenu sur tous les plans durant les dernières années. Merci pour les longues heures de re-lecture, pour les *brainstorms* d'idées et pour le soutien constant durant le long parcours menant à la réalisation de mon mémoire.

Finalement, un merci tout particulier à mes parents qui m'ont toujours encouragée dans la poursuite de mes études.

# Chapitre 1 – Introduction

Tout échange engagé entre deux individus (ou plus) est une interaction sociale (Auzanneau & Boutet, 2021; Kerbrat-Orecchioni, 1998), mais seules certaines de ces interactions correspondent à une conversation selon la définition de Sacks et al. (1974). Selon ces auteurs, une conversation est un échange verbal dont ni la longueur des tours de parole, ni l'ordre de parole, ni le contenu ne sont prédéterminés. Par conséquent, un débat n'est pas une conversation puisque le temps de parole de chaque personne, de même que le sujet de discussion et l'ordre de passage, est déjà décidé. Si les interactions sociales qui sont des conversations ne sont pas soumises aux contraintes explicites des débats, elles ne sont pas pour autant exemptes de règles. En effet, les personnes qui s'engagent dans une conversation doivent respecter un ensemble de règles implicites afin d'assurer la fluidité de ladite conversation. L'une des règles principales à la base de la communication est qu'il y a un échange du tour de parole entre les individus qui participent à la conversation : personne n'en a le monopole (Sacks et al., 1974; Stivers et al., 2009). Le partage fluide de l'espace de parole – soit la gestion des tours de parole – est un processus complexe qui reflète la coordination entre les interlocuteurs et interlocutrices et pour ainsi dire, leur capacité à interagir ensemble.

Les partenaires d'une conversation dyadique alternent entre deux rôles de manière indéterminée : celui du locuteur et celui de l'auditrice<sup>1</sup>. Si l'auditrice ne détient pas le tour de parole, son rôle n'est pas pour autant passif dans l'interaction. En effet, elle émet des signaux (vocaux, verbaux et/ou gestuels) qui contribuent au développement de la conversation (Duncan, 1972; Hömke et al., 2018; Tolins & Fox Tree, 2014; Yngve, 1970). Pour cette raison, dans la présente étude, le terme d'*interactante* sera favorisé à celui d'*auditrice* afin de mettre l'accent sur le rôle actif de la personne qui, en plus d'écouter, interagit avec le locuteur et l'influence (Kerbrat-Orecchioni, 1998).

---

<sup>1</sup> J'utiliserai la forme masculine pour désigner la personne qui parle (locuteur) et la forme féminine pour désigner la personne qui écoute (auditrice/interactante). Ce choix est strictement par souci d'inclusivité et de lisibilité du présent texte et n'a rien à voir avec une prédisposition quelconque des genres masculin et féminin à jouer un rôle particulier dans la conversation.

Les tours de parole sont influencés par le contexte dans lequel l'interaction se déroule (Boland et al., 2022; Boyle et al., 1994; Sellen, 1995; ten Bosch et al., 2005). Depuis l'avènement du téléphone et de l'ordinateur par la suite, le développement continu de la technologie offre une panoplie de nouveaux contextes d'interaction qui peuvent être utilisés à des fins récréatives et professionnelles, mais aussi en recherche. Si les rôles de locuteur et d'interactante dans la conversation sont connus depuis longtemps, les nouveaux contextes d'interaction à distance (en vidéoconférence et en réalité virtuelle) dans lesquels les partenaires de communication interagissent le sont moins et méritent d'être explorés davantage (Pan & Hamilton, 2018; Peeters, 2019). L'interaction sociale (qui passe par la communication) est au cœur du développement et du fonctionnement social. Tout au long de son développement, de l'enfance à l'âge adulte, l'individu entre en interaction avec les personnes de son entourage par la conversation, et c'est ce qui lui permet de développer son plein potentiel. Sans surprise, l'interaction sociale est associée à une meilleure santé mentale (Annahita & Mary, 2015; Ono et al., 2011; Salehi et al., 2019). Ainsi, étant donné l'importance de l'interaction sociale et considérant la grande place que les nouveaux contextes d'interaction à distance prennent (ou prendront) dans le quotidien de tous, il s'avère primordial de mieux comprendre si et comment ils agissent sur la communication.

Cela est d'autant plus vrai depuis la pandémie de COVID-19, qui a exacerbé la communication à distance. Au Québec comme ailleurs dans le monde, le confinement général imposé à la majorité des travailleurs et travailleuses a forcé l'implantation du télétravail dans beaucoup de domaines professionnels, y compris dans la pratique clinique en psychologie, en médecine et en orthophonie (voir Hamad & Jia, 2022 pour une revue récente de la littérature sur l'utilisation de la réalité virtuelle ; ou encore Ignatowicz et al., 2019 pour une revue de la littérature en lien avec l'utilisation de la vidéoconférence dans le milieu de la santé). La télépratique (soit la consultation professionnelle à distance) présente des avantages intéressants comme la possibilité d'offrir plus de services dans des régions éloignées, mais il y a un manque à combler quant à la connaissance des conséquences liées à l'utilisation des contextes de communication à distance sur le déroulement des interactions sociales. Considérant l'importance que prend l'interaction sociale dans les relations thérapeutiques, l'approfondissement des connaissances dans ce domaine est primordial à l'établissement des bonnes pratiques.

Comme nous le verrons dans les prochains chapitres, la fluidité de la conversation a un impact sur la qualité des interactions et c'est pourquoi nous souhaitons mettre en lumière les effets des nouveaux contextes d'interaction sur la coordination des partenaires. En d'autres mots, le présent mémoire a pour objectif de mieux comprendre comment la coordination entre les partenaires de conversation est influencée par les médiums de communication à distance, soit la réalité virtuelle et la vidéoconférence.

Pour atteindre cet objectif, nous présentons d'abord le contexte théorique dans lequel s'inscrit la conversation et les contextes d'interactions. Nous définissons plus précisément les concepts clés sur lesquels portent les analyses, soit les silences et les chevauchements. Nous décrivons aussi les contextes d'interaction en vidéoconférence et en réalité virtuelle en faisant le survol des études qui portent sur ces contextes d'interaction. À partir de ces définitions, nous posons la question, les objectifs et les hypothèses de cette recherche. Ensuite, nous présentons la méthode et la procédure utilisées pour répondre à la question. Nous pouvons alors rapporter les résultats détaillés de l'étude. Enfin, les résultats sont synthétisés dans une discussion approfondie visant à répondre à la question de recherche. Nous concluons ce mémoire en rappelant les faits saillants et en réfléchissant aux futurs travaux qui permettront d'enrichir les connaissances dans le domaine de la communication à distance.

## Chapitre 2 – Contexte théorique

### 2.1 L'analyse conversationnelle

Les concepts au cœur de la présente étude ont été principalement développés par les chercheurs et chercheuses qui s'inscrivent dans la lignée de l'analyse conversationnelle. Cette approche est dérivée des travaux de Harvey Sacks et ses collaborateurs (notamment dans Sacks et al., 1974). L'analyse conversationnelle est une approche qualitative qui s'intéresse aux mécanismes et à la structure des interactions verbales, et particulièrement à l'organisation de la conversation (Hepburn & Potter, 2021; Seedhouse, 2022). La méthodologie utilisée en analyse conversationnelle consiste en l'observation et la description rigoureuses de cas singuliers menant à la production de règles ou d'énoncés généraux (Kendrick, 2017). Inspirée de l'ethnométhodologie de Harold Garfinkel (1967), l'analyse conversationnelle utilise des données provenant de milieux naturels et non de l'expérimentation en laboratoire (Hepburn & Potter, 2021; Kendrick, 2017). Cette méthode de collecte de données en contexte naturel (c'est-à-dire que les interactions observées ne sont pas engendrées expressément pour l'étude en cours et que l'équipe de recherche n'influence d'aucune manière le déroulement des interactions) permet d'assurer la validité écologique des résultats. En laboratoire, il est attendu que le comportement des individus participants soit compromis par le contexte expérimental. En effet, aussitôt que les individus sont conscients qu'ils font partie d'une recherche, leurs comportements en interaction peuvent être biaisés par leurs attentes (leur désir de présenter une image positive et/ou de contribuer positivement à la recherche), par le fait qu'ils savent ce sur quoi porte l'étude (suite à la lecture du formulaire de consentement), par le fait qu'ils sont entourés de caméras et d'autre matériel expérimental visible ou encore par le fait qu'ils doivent réaliser certaines tâches plus ou moins naturelles (Kendrick, 2017).

L'analyse conversationnelle a permis le développement des connaissances dans le domaine de la communication et des interactions sociales, mais de plus en plus de chercheurs sont d'accord sur le fait que la recherche expérimentale (malgré ses défauts de validité écologique) offre une avenue intéressante pour approfondir les connaissances dans le domaine (de Ruiter & Albert,



2017; Kendrick, 2017; Seedhouse, 2022). Sans abandonner complètement l'analyse conversationnelle, on considère l'utilisation conjointe des deux méthodes (les techniques inspirées de l'analyse conversationnelle et la méthode expérimentale). Cette transgression de l'analyse conversationnelle *pure* est défendue par Kendrick (2017) qui souligne que l'opposition entre la recherche expérimentale (surtout associée à la psychologie ou à la psycholinguistique) et l'analyse conversationnelle est en fait une fausse dichotomie; les deux méthodes sont plutôt les extrémités d'un même continuum. La méthodologie utilisée pour notre étude se trouve donc au sein de ce continuum entre la méthode expérimentale et l'analyse conversationnelle. Notre collecte de données est réalisée en contexte expérimental puisque la réalité virtuelle est une nouvelle technologie qu'il est difficile de retrouver en contexte naturel. Cela dit, les données seront analysées à l'aide des concepts et des méthodes de description et d'observation de l'analyse conversationnelle. Cette méthode mixte a déjà été utilisée auparavant (par exemple, dans Finch et al., 2017; Kendrick & Holler, 2017).

## **2.2 Le tour de parole**

L'une des composantes principales de la conversation est l'échange rapide de tours de parole (de l'anglais *turn-taking*, Levinson & Torreira, 2015; Sacks et al., 1974). C'est dans les années 1970 que différents chercheurs (principalement Duncan, 1972; Sacks et al., 1974; Yngve, 1970) ont introduit le terme. Tel qu'énoncé précédemment, le droit de parole dans une conversation est partagé entre les individus qui y participent : la conversation est donc divisée en plusieurs tours de parole. Les recherches effectuées auprès de très jeunes enfants montrent la présence de tours de parole dès la période préverbale (cf. Airenti (2017) pour une synthèse de ce phénomène). La gestion des tours de parole serait d'ailleurs l'une des premières (voire la première) habileté pragmatique à se développer (voir Adams (2002) pour une synthèse du développement typique des habiletés pragmatiques).

### **2.2.1 Le système du tour de parole**

Le tour de parole est un concept central des recherches en analyse conversationnelle. En tant que système, le tour de parole organise la conversation en une alternance rapide de courts échanges verbaux. Le principe central est le suivant : les changements de tour de parole tendent à se

produire de façon fluide, soit sans silence ni chevauchement (Levinson & Torreira, 2015; Sacks et al., 1974). Le modèle de Sacks et al. (1974) propose deux composantes, (1) la construction du tour de parole et (2) l'attribution du tour de parole, ainsi qu'un ensemble de règles par lesquelles est régie la conversation.

La première composante concerne le tour de parole (la construction du tour) : le locuteur construit son tour de parole avec un ou plusieurs éléments de discours, des unités de parole (*unit-types*), qui se terminent par un « moment propice à la transition » (TRP; *transition-relevance place*). Une unité de parole peut être (sans s'y limiter) un mot, une phrase ou une courte phrase de même qu'une proposition ou un énoncé. C'est à la fin d'une unité de parole que le changement de locuteur peut, mais sans obligation, se produire. Les unités de parole seraient prévisibles de telle sorte que les autres personnes qui participent à la conversation puissent anticiper la fin du tour, soit l'arrivée d'un TRP. Cette anticipation facilite le passage fluide du tour de parole entre les interlocuteurs : l'interactante qui anticipe la venue d'un TRP est prête à prendre la parole (évitant les longs silences) et le fait au bon moment (évitant les chevauchements) (Sacks et al., 1974).

La deuxième composante du système concerne la manière dont le tour de parole passe d'une personne à l'autre, soit l'attribution du tour. Il existe deux techniques pour qu'il y ait un transfert du tour de parole à une nouvelle personne : soit la personne qui a la parole choisit le locuteur suivant (attribution par le locuteur), soit une autre personne choisit de prendre la parole (par auto-attribution) (Sacks et al., 1974).

Les composantes du système s'articulent selon un ensemble de règles, ou plutôt une série d'options, qui s'offrent lorsqu'un TRP survient à la fin d'une unité de parole. Lorsqu'un tel moment survient :

- a) Le locuteur qui a la parole peut, mais sans obligation, utiliser la technique d'attribution par le locuteur pour désigner la prochaine personne qui parlera. Dans ce cas, la personne sélectionnée a le droit, voire l'obligation, de prendre la parole.
- b) Si le locuteur actuel n'a sélectionné personne, un individu faisant partie de la conversation peut, mais sans obligation, prendre le tour de parole lui-même par auto-attribution. Si

plus d'une personne réclame le tour de parole, ce dernier revient généralement à la première personne qui s'est prononcée.

- c) Si aucune des techniques n'est utilisée, la personne qui détient le tour peut, mais sans obligation, le conserver.

Suivant cet ordre, les trois options s'alternent jusqu'à ce qu'une personne prenne effectivement la parole.

Depuis le modèle élaboré par Sacks et ses collaborateurs, plusieurs recherches appuient l'existence d'une organisation de la conversation en tours de parole. En effet, l'organisation de la conversation en tours de parole serait une composante universelle du langage. Une étude portant sur le tour de parole auprès d'un échantillon regroupant 10 langues différentes montre la présence d'une organisation de la conversation en tours de parole dans toutes les langues étudiées (Stivers et al., 2009). D'autant plus, ces travaux montrent que, dans toutes les langues étudiées, le temps de réponse pour une question fermée (de type oui/non) est majoritairement d'environ 200 millisecondes (ms). Ces délais de réponse très rapides ont aussi été rapportés dans des discussions spontanées (Bögels, 2020; Heldner & Eklund, 2010). Ces transitions rapides entre les partenaires de communications appuient le principe des changements de tour de parole « sans silence ni chevauchement ».

Par ailleurs, d'autres études qui se penchent sur le rôle de l'anticipation de l'interactante dans la réalisation des échanges de tours de parole fluides appuient aussi le modèle de Sacks et ses collaborateurs (Magyari & de Ruiter, 2012; Riest et al., 2015). Pour planifier une réponse courte, comme nommer un objet simple, une personne a besoin d'environ 600 ms (Indefrey & Levelt, 2004). Ce délai de planification augmente à minimalement 1 000 ms pour la description d'un événement (Konopka & Meyer, 2014). Ce faisant, pour émettre une réponse avec un délai de seulement 200 ms, l'interactante doit planifier sa réponse avant même que le locuteur ait terminé son tour de parole, tout en traitant simultanément l'information qu'elle reçoit (Barthel & Levinson, 2020; Corps et al., 2018). La prise de tour de parole fluide est donc une activité complexe qui requiert simultanément le traitement de l'information reçue, la planification de la réponse et l'anticipation du TRP afin d'émettre la réponse au bon moment.

### **2.2.2 Les facteurs qui influencent le tour de parole**

La prise de parole sans silence ni chevauchement est influencée par différentes caractéristiques de l'interaction sociale qui dépassent l'échange d'informations verbales. Premièrement, la familiarité entre les partenaires de communication influencera la manière dont ils interagissent entre eux. Par exemple, les partenaires de communication qui ne se connaissent pas feraient moins de chevauchements dans leur conversation comparativement à ceux qui se connaissent (Bortfeld et al., 2001).

Deuxièmement, la gestion des tours de parole est influencée par la communication non-verbale : entres autres les gestes, les regards et les expressions faciales. La gesticulation du locuteur, qui accompagne la parole et même les silences, indique à l'interactante s'il conserve son tour de parole (gesticulation continue) ou s'il l'achève (fin de la gesticulation) (Duncan, 1972; Duncan & Fiske, 1977). Par ailleurs, des phénomènes aussi subtiles que le clignement des yeux de l'interactante peuvent influencer le discours du locuteur : lorsque le clignement des yeux de l'interactante est plus long, le locuteur parle moins longtemps que lorsque ses clignements sont plus courts (Hömke et al., 2018). Si le regard de l'interactante influence le locuteur, le contraire est aussi vrai : le regard du locuteur influence à son tour l'interactante. Par exemple, celle-ci répond plus rapidement lorsque le locuteur la regarde en posant sa question que s'il ne la regarde pas (Stivers et al., 2009). Par conséquent, le contact visuel joue un rôle dans la gestion des tours de parole (cf. Degutyte & Astell, 2021; Zima et al., 2019, pour une synthèse des fonctions du regard). De plus, la communication non-verbale (par exemple, le hochement de la tête) joue un rôle important dans la gestion des tours de parole puisqu'elle peut être présente en l'absence de parole (durant les silences). En effet, si l'interactante hoche la tête durant un silence qui suit un TRP, celle-ci maintien la communication avec le locuteur en lui indiquant, par exemple, qu'elle souhaite lui laisser la parole (Duncan, 1972). Malgré le silence, la coordination entre les partenaires semble fluide grâce au hochement de tête qui permet de maintenir l'interaction (Kogure, 2007). L'interaction pendant les silences peut être maintenue grâce aux hochements de la tête, mais aussi grâce à d'autres formes de communication non-verbale qui peuvent se produire durant les silences comme les expressions faciales ou encore grâce à des gestes qui remplacent le langage comme faire « non » de tête ou hausser les épaules (Cosnier, 2016).

Finalement, le contexte d'interaction dans lequel se déroule la conversation est aussi un facteur d'influence important de la coordination entre les partenaires de conversation. La distinction entre les interactions au téléphone (ou plutôt les interactions par voie audio seulement) et en face-à-face n'est plus à faire (Boyle et al., 1994; Doherty-Sneddon et al., 1997; ten Bosch et al., 2005). Cela dit, les nouvelles technologies offrent de nouveaux médiums de communication permettant la transmission instantanée d'information audio et visuelle comme la vidéoconférence et la réalité virtuelle. Malheureusement, la littérature scientifique sur le sujet est relativement pauvre.

L'interaction sociale est au cœur du développement et du fonctionnement social. Les silences et les chevauchements peuvent avoir une influence sur la qualité de l'interaction sociale<sup>2</sup> puisqu'ils sont associés à plus d'émotions négatives et à moins d'engagement (Koudenburg et al., 2013). Puisque de plus en plus d'interactions passent maintenant par des médiums de communication à distance, nous souhaitons répondre au manque de connaissances actuelles en ce qui a trait aux effets de la vidéoconférence et de la réalité virtuelle sur la fluidité de la conversation, qui elle-même influence la qualité de l'interaction sociale. Pour ce faire, les sections suivantes approfondissent d'abord les indices de la coordination des tours de parole : les silences et les chevauchements. Nous faisons ensuite l'état des connaissances quant aux contextes d'interaction en vidéoconférence et en réalité virtuelle.

## **2.3 La coordination entre les partenaires de communication**

Les silences et les chevauchements sont des indicateurs de la coordination entre les participants et participantes de la conversation : dans une conversation dite fluide, les échanges se déroulent sans silence ni chevauchement (Sacks et al., 1974). Ainsi, les individus organisent la conversation par l'échange rapide de tours de parole durant lesquels essentiellement une seule personne parle à la fois (Levinson & Torreira, 2015).

---

<sup>2</sup> La qualité de l'interaction sociale, telle qu'utilisée ici, fait référence à la manière dont l'interaction est vécue par les personnes qui y participent. Une interaction sociale d'une bonne qualité est normalement associée à des sentiments positifs. Par exemple dans Monninger et al. (2023), la qualité de l'interaction sociale est mesurée par auto-évaluation de l'expérience d'interaction sociale sur une échelle de « très négative » à « très positive ».

### 2.3.1 Les silences

Dans une conversation, un silence est un moment durant lequel personne ne parle (Levinson & Torreira, 2015). Quoiqu'il soit par définition une absence de parole, le silence n'est pas nécessairement synonyme d'absence de communication. Ephratt (2008) distingue trois types de silences : l'inertie (*stillness*), les pauses et les silences éloquents. L'inertie fait référence aux situations dans lesquelles il n'y a pas d'interaction. Par exemple, lorsque deux inconnus se croisent dans la rue sans se parler, il y a une absence de parole qui résulte du fait que ces personnes n'entrent pas en communication : il y a inertie. Quant aux pauses, ce sont des silences qui sont causés par une suite d'événements. Par exemple, une personne qui prend une pause pour respirer avant de (re)commencer à parler. Les pauses, de même que l'inertie, ne sont pas des silences qui ont une fonction communicationnelle. Ce sont les silences éloquents qui sont porteurs de sens dans une situation de communication. Ephratt (2008) subdivise ce type de silences selon six fonctions du langage de Jakobson (1960) incluant par exemple la fonction expressive, phatique et poétique. La fonction des silences qui est d'intérêt pour la présente recherche est celle de marqueur du discours (qui fait partie de la fonction conative). Bien que Sacks et al. (1974) n'aient jamais utilisé explicitement le terme « marqueur de discours » c'est dans cette catégorie que se trouvent les silences décrits dans leurs travaux sur le tour de parole (Ephratt, 2008). Ce sont ces silences qui témoignent de la fluidité de la conversation, soit la capacité des interlocuteurs à prendre la parole au moment opportun.

En analyse conversationnelle, trois catégories de silences se distinguent en fonction du contexte dans lequel le silence se produit : les pauses, les intervalles (*gaps*) et les laps (Sacks et al., 1974). D'un côté, les pauses sont des silences qui surviennent à l'intérieur d'un même tour de parole. Ce sont des silences intra-locuteurs qui n'impliquent pas de changement de locuteur. À l'inverse, les intervalles sont des silences qui surviennent entre deux locuteurs. Ce sont des silences inter-locuteurs à la suite desquels le tour de parole passe à un nouveau locuteur. Finalement, les laps sont de longs silences qui se situent à un TRP. Plus précisément, les laps surviennent à un TRP lorsque : *a*) le locuteur actuel n'attribue pas le tour de parole ; *b*) personne ne le réclame par auto-attribution et *c*) le locuteur ne poursuit pas son tour. Dans ce cas, le modèle de Sacks et al. (1974) prévoit qu'il y ait un enchaînement en boucle des règles d'attribution du tour de parole

jusqu'à ce qu'une personne prenne finalement la parole. C'est ce phénomène qui explique la durée plus longue des silences à certains TRP, soit les laps. Le processus d'enchaînement en cascade des règles d'attribution du tour de parole explique aussi pourquoi les pauses sont généralement plus longues que les intervalles : ce n'est qu'en dernier recours, si le locuteur n'attribue pas le tour de parole et si personne ne le réclame que le locuteur courant reprend la parole. Comme d'autres chercheurs et chercheuses auparavant (par exemple, Chowdhury et al., 2017; Heldner & Edlund, 2010; Levinson & Torreira, 2015), ces catégories (les pauses intra-locuteurs, les intervalles inter-locuteurs et les laps) serviront à la catégorisation des silences observés dans la présente étude.

Pour Sacks et al. (1974), ce sont les silences les plus longs – les laps – qui sont responsables des discontinuités dans les conversations. En d'autres mots, la présence de silences dans la conversation ne serait pas nécessairement synonyme d'un manque de fluidité, à condition que ces silences soient de courtes durées. Dans le même ordre d'idées, les silences ont normalement une durée maximale moyenne d'une seconde (Jefferson, 1989). Au-delà de ce délai, les partenaires de communication auraient tendance à intervenir pour briser le silence. Cet évitement des longs silences pourrait s'expliquer par le fait que ces silences sont associés à des émotions négatives. Plus précisément, une étude montre que les conversations qui contiennent des silences d'une seconde sont associées à des émotions négatives et à moins d'engagement que celles qui sont fluides (Koudenburg et al., 2013). De plus, la longueur des silences influence la manière dont les propos des locuteurs sont perçus. En effet, lorsqu'un changement de tour de parole survient avec un long silence inter-locuteur (un long intervalle), le locuteur qui prend la parole est perçu comme étant moins honnête que lorsque l'intervalle est plus court (Fox Tree, 2002). Il n'y a pas que la durée des silences qui affecte la fluidité de la conversation, mais aussi la fréquence de ceux-ci. En effet, les conversations les plus fluides sont celles qui ont le moins grand nombre de silences (Lestary et al., 2018). Par conséquent, les silences indiquent que la conversation manque de fluidité lorsqu'ils sont trop longs et/ou trop nombreux.

### 2.3.2 Les chevauchements

Les chevauchements sont les moments durant lesquels plus d'une personne parle en même temps. Dans certains contextes, les paroles simultanées ne sont pas problématiques. Par exemple, il est attendu que plusieurs personnes s'expriment en même temps lors d'activités conjointes comme dans une chorale, pour une prière ou encore lors de salutations (Schegloff, 2001). Par conséquent, il faut distinguer les différents types de parole simultanée afin de cibler, dans le cas présent, les paroles simultanées qui sont des chevauchements dus à des problèmes de coordination entre les partenaires de communication. Plus précisément, nous distinguerons les régulateurs<sup>3</sup> (qu'on appellera par la suite simplement « BC » en référence au terme *backchannel* en anglais) et les interruptions de ces chevauchements.

Tout d'abord, les BC sont les chevauchements qui sont les plus fréquents (Levinson & Torreira, 2015). Quoique la définition du BC diverge d'un auteur à l'autre, au moins trois éléments semblent communs. (1) Le BC est un comportement de l'interactante qui se produit pendant le tour de parole du locuteur sans l'interrompre et (2) il ne constitue pas une prise de parole. De plus, (3) le comportement peut être vocal, verbal ou gestuel (par exemple, Bangerter & Clark, 2003; Duncan, 1972; Knudsen et al., 2020; Tolins & Fox Tree, 2014; Xu, 2014; Yngve, 1970). Le plus souvent, les BC vocaux ou verbaux sont de très courtes interventions comme « oui » et « hu-hum » (Knutson, 2009). Comme les BC sont par définition des intrusions qui ne mènent pas à l'interruption du discours du locuteur, ces interventions ne sont pas des cas de chevauchement qui témoignent des problèmes de coordination dans la conversation en ce qu'ils ne mènent pas à l'enclenchement de mécanismes de résolution (Schegloff, 2001).

Ensuite, des paroles simultanées peuvent se produire lors d'interruptions. De façon générale, les interruptions font référence aux moments où l'interactante débute délibérément un énoncé verbal (qu'elle complètera ou non) au milieu du tour de parole du locuteur (Schegloff, 2001). En

---

<sup>3</sup> L'activité de l'interactante porte différents noms dans la littérature : il est parfois question de *backchannels* (Duncan, 1972; Yngve, 1970), de *listener responses* (Dittmann & Llewellyn, 1968; Eiswirth, 2020; Ross, 2018), de *receptive tokens* (Clancy et al., 1996) ou encore de *feedback* (Kraut et al., 1982; Wagner et al., 2014). À l'instar de Kerbrat-Orecchioni (1998) et Cosnier (2016), le terme « régulateur » a été choisi pour référer au *backchannel* (BC) qui englobe généralement l'ensemble des comportements de l'interactante qui sont produits parallèlement, en réaction au discours principal du locuteur.



analyse conversationnelle, l'une des principales distinctions entre les interruptions et les chevauchements est leur contexte d'apparition : les interruptions surviennent dans le discours du locuteur en l'absence d'un TRP. À l'inverse, les chevauchements se produisent généralement à un moment où la fin du tour du locuteur est raisonnablement envisageable (Hepburn & Potter, 2021).

Finalement, les paroles simultanées qui sont associées au manque de fluidité dans la conversation sont généralement des chevauchements qui résultent de deux événements simultanés : l'interactante prend la parole puisqu'elle anticipe raisonnablement la fin du tour de parole du locuteur alors qu'au même moment, le locuteur poursuit, dans ses droits, son tour de parole (Jefferson, 1986). Tel que mentionné, ces chevauchements surviennent donc à des TRPs, soit à la suite d'un silence lorsque les deux partenaires de communication réclament le tour de parole en même temps ou encore lorsque l'interactante anticipe mal la fin du tour du locuteur et commence son tour de parole avant la fin de ce tour.

De la même manière que les silences, si les chevauchements sont généralement évités dans la conversation, ils en sont rarement complètement absents. Cela étant dit, ce ne sont pas tous les chevauchements qui semblent poser un problème dans la conversation. Pour cette raison, nous porterons une attention particulière aux chevauchements qui occasionnent des bris de communication n'impliquant ni de BC ni d'interruptions délibérées. Un bris de communication est un moment durant lequel il y a un problème dans la conversation (causé par une mauvaise production, compréhension ou réception d'un énoncé) qui implique une coupure dans la fluidité de la conversation et par conséquent, qui doit être « réparé ». Le bris de communication comprend quatre phases (tel que décrit dans Orange et al., 1998). Premièrement, il y a la source du problème, soit l'événement qui cause le bris, l'interruption dans la fluidité. Les sources de problème sont très diverses : ce peut être essentiellement n'importe quel événement qui déclenche un mécanisme de réparation. Deuxièmement, il y a la reconnaissance du problème. Les interlocuteurs qui reconnaissent le problème émettent des signaux pour signifier l'un à l'autre qu'une réparation est nécessaire. Troisièmement il y a la phase de réparation durant laquelle les interlocuteurs tentent de réparer le bris de communication. Finalement, la quatrième phase réfère au résultat de la réparation, soit la résolution du bris de communication.

Ainsi, un bris de communication, comme un chevauchement, provoque le déploiement de mécanismes de réparation permettant la reprise fluide de la conversation (Hayashi et al., 2013). Depuis les travaux importants de Emanuel A. Schegloff et ses collaborateurs (1977), on considère deux types de réparation : les réparations auto-initiées par la personne ayant causé la source du bris et les réparations initiées par l'autre (soit l'interactante). Ces réparations se distinguent en fonction de l'endroit où elles apparaissent et en fonction des différentes techniques d'initiation utilisées. D'un côté, les réparations auto-initiées se trouvent principalement à trois emplacements : dans le même tour que la source du bris, à l'intérieur du TRP ou encore dans le tour suivant la source du trouble. D'un autre côté, les réparations initiées par l'autre se trouvent généralement à un seul endroit, soit dans le tour suivant la source du problème (Schegloff et al., 1977). Finalement, les réparation auto-initiées peuvent être initiées par une rupture du discours, l'allongement d'un son ou encore un « uh » alors que les réparations initiées par l'autre sont plutôt des questions qui demandent une clarification comme « quoi ? », « quand ? » et « qui ? ».

En conclusion, la présence de silences et de chevauchements dans les conversations est à la fois normale, voire inévitable, mais aussi signe de perturbations. Nous nous intéresserons aux silences qui sont particulièrement longs (relativement aux autres) et aux chevauchements qui impliquent un bris dans la conversation (nécessitant une réparation), phénomènes qui sont davantage associés au manque de fluidité.

## **2.4 Les contextes d'interaction**

L'intérêt pour la communication à distance a connu un essor remarquable dans les dernières années, surtout en raison de la pandémie de COVID-19 et du confinement à l'échelle mondiale. La présente étude se concentre sur les contextes en vidéoconférence et en réalité virtuelle puisque ce sont des technologies récentes qui, contrairement au téléphone, permettent la transmission de l'information audio et visuelle.

### **2.4.1 La vidéoconférence**

La vidéoconférence est un moyen de communication via un ordinateur qui permet la communication par la transmission audio et vidéo simultanée entre deux personnes ou plus. Si

l'accès à l'information visuelle et audio semble a priori faire de la vidéoconférence un contexte d'interaction semblable à l'interaction en face-à-face, les recherches dans le domaine ne sont pas concluantes. Déjà en 1997, une première étude utilisant la tâche de dessin de carte (*Map Task*) s'intéresse à l'interaction par vidéo en comparaison à l'interaction en face-à-face (Doherty-Sneddon et al., 1997). La tâche de conversation implique deux partenaires qui doivent communiquer oralement pour reproduire une image correctement : une personne a la solution sous les yeux et doit guider l'autre personne qui doit placer des pièces aux bons endroits sur sa propre carte. Les résultats de l'étude montrent que, pour un même niveau de performance à la tâche de communication, les dyades utilisent plus d'énoncés (elles parlent plus) lorsqu'elles sont dans le contexte en vidéo que dans celui en face-à-face (Doherty-Sneddon et al., 1997). Si les procédures utilisées pour créer le contexte d'interaction par vidéo dans cette étude sont loin de la vidéoconférence d'aujourd'hui, les résultats indiquent à tout le moins que le contexte d'interaction influence les interactions et que l'accès à l'information sonore et visuelle n'est pas suffisant pour créer une interaction aussi authentique que celles en face-à-face. Par ailleurs, une étude plus récente sur les interactions en vidéoconférence souligne que dans ce contexte, les personnes ajustent leurs comportements : elles auraient tendance à exagérer leurs comportements non-verbaux et à parler plus fort que lors d'une interaction en face-à-face (Croes et al., 2019). Deux caractéristiques de la vidéoconférence semblent influencer plus particulièrement la coordination entre les partenaires de conversation : les délais créés par le traitement de l'information et les indices visuels disponibles.

En premier lieu, dans les interactions qui sont médiées par des logiciels vidéos (voire par toute forme de technologie), les individus n'ont pas accès aux informations visuelles et verbales en temps réel : il y a de la latence. La latence correspond au temps nécessaire pour que l'information soit transmise d'une personne à l'autre.<sup>4</sup> En effet, entre le moment où une personne émet un comportement et la perception de ce comportement par l'autre personne, l'information visuelle et sonore doit être traitée et transformée par plusieurs intermédiaires technologiques (entre

---

<sup>4</sup> Il est difficile d'identifier une durée moyenne de la latence puisqu'elle varie en fonction de divers éléments, entre autres la qualité des appareils utilisés ainsi que la vitesse et l'achalandage du réseau Internet. À titre d'exemple, Boland et al. (2022) rapportent des latences en Zoom qui varient entre 30 et 70 ms.

autres la caméra, l'ordinateur, le logiciel de vidéoconférence et Internet). D'un côté, selon l'étude de Seuren et al. (2021), une latence d'environ 100 ms est peu problématique alors qu'une latence de 700 ms influence la manière dont les tours de parole s'organisent et rend la coordination entre les personnes plus complexe. D'abord, même si un locuteur initie son tour de parole dans un temps raisonnable, l'interactante pourrait percevoir un délai de réponse plus long (un silence) à cause du temps de latence. Ensuite, étant donné ce décalage entre les partenaires de communication, il arrive que ces derniers ne perçoivent pas le début d'un chevauchement au même moment. De ce fait, la résolution dudit chevauchement est retardée puisqu'il s'écoulera plus de temps avant que les partenaires mettent en place les stratégies de résolution permettant de réparer le chevauchement (Seuren et al., 2021). De l'autre côté, Boland et al. (2022) suggèrent qu'une latence plus petite (aussi peu que 30 à 70 ms) puisse perturber la fluidité de la conversation. Les résultats de leurs expériences montrent des délais de réponse en vidéoconférence (en moyenne 976 ms pour des questions de type oui/non) supérieurs aux délais de latence moyens (30-70 ms). Autrement dit, la latence causée par le traitement de l'information à distance n'expliquerait pas, à lui seul, les délais plus longs en vidéoconférence. Les auteurs suggèrent que la variabilité des délais de latence (leur caractère imprévisible) a un effet sur les mécanismes de synchronisation automatique neuronale (oscillateurs)<sup>5</sup> des partenaires de communication. Ce serait ce dérèglement dans la synchronisation qui serait à l'origine des longs délais observés en vidéoconférence.

En deuxième lieu, si la vidéoconférence permet de transmettre de l'information visuelle, cette information est plus limitée qu'en face-à-face. D'une part, l'information visuelle qui est transmise est limitée par le cadre de la caméra. Généralement, les personnes sont assises devant l'écran d'ordinateur et la caméra est placée de sorte que le visage soit au centre de l'image transmise. De cette manière, l'information visuelle en vidéoconférence ne contient que le haut du corps (de la tête jusqu'aux épaules). Ainsi, la vidéoconférence ne peut pas capter toutes les informations

---

<sup>5</sup> Pour résumer simplement, le mécanisme d'oscillateurs neuronales vient principalement du travail de Wilson et Wilson (2005) qui suggèrent que les partenaires de conversation utilisent le rythme de la parole (le taux de syllabes prononcées par seconde) pour synchroniser leur interaction et prévoir leurs tours de parole. Nous suggérons de consulter l'article originale de Wilson et Wilson (2005) ou la synthèse de Boland et al. (2022) pour une compréhension plus exhaustive du modèle.

paralinguistiques, entre autres le mouvement du corps et la gesticulation des mains, qui sont normalement en jeu dans une conversation en face-à-face (Bohannon et al., 2013). La vidéoconférence affecte aussi la manière dont le regard est transmis. Si une personne veut donner l'impression de regarder directement l'autre, elle doit regarder la caméra et non le moniteur sur lequel est diffusée l'image de son partenaire de conversation. Ce faisant, il est impossible de reproduire le même type de contact visuel en vidéoconférence qu'en face-à-face puisqu'il est impossible de fixer la caméra et les yeux de l'autre en même temps (Bohannon et al., 2013). Comme le contact visuel a un effet sur la coordination des tours de parole (Degutyte & Astell, 2021), cette coordination risque d'être altérée en contexte de vidéoconférence.

Finalement, la vidéoconférence est au cœur de nombreuses études récentes qui portent sur l'efficacité de traitements et de thérapies à distance. Selon ces recherches, la vidéoconférence n'est pas appropriée à toutes les clientèles (Ignatowicz et al., 2019; Schutz et al., 2022). Par conséquent, il s'avère important de mieux comprendre les effets de la vidéoconférence sur la communication, afin de mieux cibler les clientèles pour qui ce média peut être adéquat.

#### **2.4.2 La réalité virtuelle**

La réalité virtuelle rejoint la vidéoconférence dans la catégorie des moyens de communication à distance qui permettent de transmettre des informations visuelles et audio. La réalité virtuelle réfère à un monde virtuel tridimensionnel, créé par un ordinateur, dans lequel la personne est en immersion complète (Peeters, 2019). Elle offre aux chercheurs et chercheuses la possibilité de créer un environnement expérimental contrôlé et répliquable, tout en ayant une grande validité écologique (Pan & Hamilton, 2018). Dans le domaine de la psychologie, la réalité virtuelle montre une efficacité prometteuse dans le traitement des phobies : elle permet d'exposer la personne à sa phobie dans un environnement naturel mais contrôlé (Bouchard et al., 2007; Malbos et al., 2013). Une revue récente portant sur l'utilisation de la réalité virtuelle dans le traitement de divers troubles en santé mentale mentionne que l'exposition en réalité virtuelle est efficace pour le traitement de phobies spécifiques, mais aussi dans le traitement de l'anxiété et du trouble de stress post-traumatique (Park et al., 2019).

Cependant, la réalité virtuelle n'a pas que des avantages : son utilisation peut impliquer différents problèmes. Par exemple, il peut y avoir une discordance entre les mouvements réellement produits par un individu et la manière dont ceux-ci sont transmis dans l'environnement virtuel (de la Rosa & Breidt, 2018). De plus, les interactions en réalité virtuelle se font par le biais d'avatars, des personnages animés virtuels qui peuvent avoir une forme plus ou moins humanoïde. Donc, contrairement à la vidéoconférence, ce n'est pas l'image réelle de la personne qui est transmise. L'utilisation d'avatars en réalité virtuelle implique que l'information visuelle transmise lors de l'interaction est limitée à des mouvements grossiers (mouvement de la tête, des mains et du corps) puisque la technologie ne permet pas aux avatars de transmettre des informations plus subtiles comme la direction du regard, les sourires ou toutes autres expressions faciales normalement utilisées dans les interactions sociales (Pan & Hamilton, 2018; Peeters, 2019). Par conséquent, les personnes qui interagissent en réalité virtuelle n'ont pas accès aux mêmes informations qu'en vidéoconférence pour assurer leur coordination durant leurs échanges.

Les études qui portent sur la réalité virtuelle se concentrent souvent sur l'efficacité des thérapies en santé mentale et même en réadaptation (Lu et al., 2022; Park et al., 2019). À cet effet, la réalité virtuelle est utilisée comme un outil de thérapie où l'environnement virtuel est contrôlé « parfaitement » par l'équipe de recherche dans un but précis. Par exemple, dans l'étude de Bouchard et al. (2017), l'environnement virtuel est manipulé afin d'exposer le patient à des scénarios particuliers dans le but d'évaluer l'efficacité d'une thérapie d'exposition en réalité virtuelle. Dans leurs scénarios sociaux en réalité virtuelle, les avatars qui interagissent avec le patient sont contrôlés par le thérapeute, mais ne reproduisent que des réponses pré-programmées. Ce faisant, l'avatar n'est pas *incarné* par le thérapeute, mais il est plutôt *commandé* par ce dernier ; l'avatar ne reproduit pas les faits et gestes du thérapeute, il exécute les comportements que le thérapeute le charge de faire. Ainsi, les études sur l'efficacité des thérapies en réalité virtuelle informent sur la manière dont les participants perçoivent l'environnement virtuel, mais elles ne mettent pas en lumière la manière dont les comportements sont transposés (ou plutôt reproduits) dans l'environnement virtuel. Dans la perspective où la réalité virtuelle pourrait être utilisée comme pur moyen de communication entre deux personnes

(par exemple, en thérapie pour faire de la télépratique), il est nécessaire de bien comprendre comment les divers contextes à distance permettent (ou non) au thérapeute (ou autre professionnel) de transmettre ses comportements durant une thérapie. Il est primordial pour les professionnels de bien comprendre les effets et les limites de la réalité virtuelle sur les comportements qu'ils peuvent projeter à leur patient en interaction, mais aussi de reconnaître les conséquences de la réalité virtuelle sur les comportements perceptibles (ou non) de leur patient. C'est pourquoi le présent mémoire s'intéresse à la réalité virtuelle en tant que contexte d'interaction où deux personnes sont en immersion complète.

L'approfondissement des connaissances sur les effets des différents modes de communication à distance sur l'interaction sociale permettra d'envisager un usage en télépratique qui est avisé pour les différents types de clientèle. À notre connaissance, puisque la réalité virtuelle est une nouvelle technologie de communication, peu d'études comparent les deux contextes d'interaction l'un à l'autre et aucune ne s'intéresse à leur effet sur la coordination entre les partenaires de conversation<sup>6</sup>. Pourtant, la communication à distance (plus précisément la télépratique) devient une alternative de plus en plus populaire à l'interaction en personne, ce qui rend d'autant plus importante la tâche d'éclairer les enjeux propres aux différents contextes d'interaction.

---

<sup>6</sup> Nous avons trouvé une étude publiée très récemment qui compare les deux contextes d'interaction, mais elle s'intéresse au sentiment d'engagement et de présence et non à la coordination (Gruber et al., 2023).

## Chapitre 3 – Objectifs et hypothèses

Selon le principe « sans silence ni chevauchement » (ou la réduction au minimum de ces phénomènes), il y a un manque de fluidité dans les échanges de tour de parole lorsque le prochain locuteur (1) prend la parole trop tard ou encore (2) prend la parole trop tôt. En d'autres mots, la fluidité de la conversation est compromise par la présence de silences et de chevauchements.

Le manque de fluidité dans la conversation a un impact négatif sur l'interaction sociale. D'une part, la présence de longs silences est associée à plus d'émotions négatives et à des sentiments de rejet alors que les conversations qui sont fluides (sans silences) sont associées à plus de sentiments d'appartenance, d'estime de soi et de validation sociale (Koudenburg et al., 2011). D'autre part, la présence de chevauchements implique des bris dans la communication qui peuvent provoquer des problèmes de compréhension entre les partenaires de conversation. La présence fréquente de chevauchements peut donc engendrer de la frustration dans l'interaction sociale. Ainsi, la capacité à échanger des tours de parole de façon fluide est liée à la capacité à socialiser et à se développer (Khademi, 2022).

Tel qu'annoncé en introduction, le but de cette étude est donc de répondre à la question suivante : en quoi la coordination en réalité virtuelle est-elle différente de celle en vidéoconférence ?

### 3.1 Premier objectif : analyse quantitative

Pour répondre à la question de recherche, le premier objectif de ce mémoire est de comparer la coordination dans les différents contextes d'interaction à l'aide de données quantitatives. Plus précisément, nous cherchons à savoir si les silences et les chevauchements qui ont été décrits dans le contexte théorique comme les marqueurs de la coordination sont différents d'un contexte à l'autre en termes de fréquence et de durée.

D'une part, la VR pourrait nécessiter un traitement informatique plus complexe que la vidéoconférence. Pour transmettre les mouvements d'une personne dans le monde virtuel, l'ordinateur doit traiter l'information spatio-temporelle provenant des manettes et du casque de



VR, et reproduire ces informations par le biais d'un avatar dans un environnement virtuel. Ce traitement plus complexe risque de créer plus de latence, ce qui rend la tâche d'assurer la fluidité de la conversation plus difficile pour les partenaires. D'autre part, étant donné qu'en VR, les partenaires n'ont pas accès à l'information visuelle subtile comme les expressions faciales et la direction du regard (Pan & Hamilton, 2018; Peeters, 2019), les échanges en réalité virtuelle devraient être moins fluides que ceux en vidéoconférence. Pour ces deux raisons, les silences en VR devraient être plus nombreux (H1) et plus longs (H2) que ceux en vidéoconférence. De même, les chevauchements devraient être plus fréquents en VR (H3). Suivant le principe de Sacks et al. (1974), les conversations qui sont moins fluides (ou plutôt durant lesquelles les partenaires ont plus de difficultés à coordonner leurs échanges de tours de parole) s'effectuent avec plus de silences et de chevauchements. Donc, si nos hypothèses sont validées, nous pourrions conclure que le contexte en VR influence plus négativement la coordination entre les partenaires durant la conversation que le contexte en vidéoconférence.

### **3.2 Second objectif : analyse qualitative**

Le deuxième objectif de ce mémoire est de comparer les moments qui sont problématiques (soit les moments qui contreviennent au principe « sans silence ni chevauchement ») dans les différents contextes d'interaction. Suivant les principes de l'analyse qualitative, les descriptions des interactions lors de silences et de chevauchements seront comparées d'un contexte à l'autre. Cette analyse permettra de mettre en lumière les similitudes et/ou les différences entre les contextes d'interactions en ce qui a trait aux problèmes de coordination dans la conversation. Étant donné que la réalité virtuelle est une technologie récente et que son étude en tant que mode de communication est émergente, notre démarche est exploratoire. Pour cette raison, il ne nous semble pas raisonnable (voir même possible) d'émettre des hypothèses quant aux phénomènes qui seront mis en lumière par notre analyse.

Si ces deux objectifs sont atteints, nous saurons quelles sont les différences quantitatives de la coordination dans les contextes d'interaction à distance, mais aussi quels éléments qualitatifs sont en jeu dans chacun des contextes lorsque la conversation manque de fluidité, ce qui nous permettra de relever certaines caractéristiques propres à chacun des contextes.

## **Chapitre 4 – Méthodologie**

Cette section décrit la méthode par laquelle les deux objectifs permettant de répondre à la question de recherche du présent mémoire sont atteints.

### **4.1 Échantillonnage**

La population cible de cette étude est la population de jeunes adultes québécois dont la langue maternelle est le français. Plus précisément, les personnes doivent être âgées entre 18 et 30 ans et n'avoir aucun trouble développemental du langage, aucun trouble du comportement (par exemple, un trouble de la personnalité opposante) ni toute autre condition biomédicale connue (par exemple, un trouble du spectre de l'autisme, trisomie 21, etc.). Les individus doivent avoir été scolarisés au Québec et leur langue maternelle doit être le français. De plus, les deux partenaires de communication doivent se connaître depuis au moins trois ans puisque les interactions sociales sont influencées par le fait que les interlocuteurs soient connus ou inconnus l'un de l'autre (McFarland, 2001).

L'échantillon est créé par échantillonnage intentionnel – selon les critères d'inclusion et d'exclusion – pour contrôler l'influence des phénomènes qui peuvent agir sur le déroulement des échanges, soit l'âge des partenaires de conversation, le fait qu'ils se connaissent ou non et la présence d'un trouble qui influence les compétences interactionnelles. Le recrutement de dyades qui respectent ces critères nous assure d'avoir un échantillon présentant une certaine homogénéité.

### **4.2 Éthique**

Une approbation éthique (Projet CEREP-21-148-D) a été octroyée par le Comité d'éthique de la recherche en éducation et en psychologie (CEREP) de l'Université de Montréal pour le recrutement des participants et participantes pour le projet principal et pour l'analyse des données pour la présente recherche. Les individus ayant participé à la recherche ont été informés de l'enregistrement audio et vidéo de leurs interactions et y ont consenti avant de procéder à l'expérimentation.

### 4.3 Procédure

Les deux partenaires de communication sont invités à un entretien d'environ 120 minutes au laboratoire de recherche en compagnie d'un membre de l'équipe de recherche. Un court questionnaire portant sur les habitudes d'utilisation des technologies de communication ainsi que sur les informations socio-démographiques est complété préalablement à la rencontre. Durant la visite au laboratoire, les partenaires de conversation doivent s'entretenir dans un total de cinq contextes d'interaction différents : en présence, au téléphone, en réalité virtuelle sans casque, en réalité virtuelle avec le casque et en vidéoconférence. La présente étude n'utilise que les interactions dans les contextes en présence, en vidéoconférence et en réalité virtuelle avec le casque. Pour le contexte en présence, les individus sont assis l'un en face l'autre à une distance d'environ deux mètres (pour respecter les mesures sanitaires conséquentes à la pandémie). Pour les contextes à distance (vidéoconférence et VR), les deux partenaires de conversation sont séparés dans des salles individuelles et ne peuvent communiquer que par l'entremise d'un ordinateur de l'équipe de recherche. Pour la situation en vidéoconférence, les individus sont assis devant l'ordinateur et peuvent se voir (de la tête aux épaules) et s'entendre mutuellement via le logiciel Zoom. Pour la situation en réalité virtuelle, les individus se tiennent debout au centre d'un espace de jeu (2x2 mètres environ). Grâce au casque de réalité virtuelle Oculus Rift-S, ils sont en immersion complète dans l'environnement tridimensionnel du logiciel VRChat dans lequel ils peuvent interagir par le biais d'avatars.

Pour chaque contexte d'interaction, un troisième ordinateur (installé dans une troisième salle) permet à l'équipe de recherche d'assister aux échanges en observation non-participante. Tous les échanges sont enregistrés par ce troisième ordinateur. De plus, des caméras indépendantes installées dans chaque salle permettent de capter l'environnement physique dans lequel se trouvent les participantes et participants.

Les contextes d'interaction sont présentés dans un ordre aléatoire à chacune des dyades de sorte que le contexte en vidéoconférence ne soit pas toujours présenté avant (ni toujours après) le contexte en réalité virtuelle, par exemple. Cette procédure a été choisie afin de contrôler pour

les effets de nouveauté et d'habituation des participantes et participants quant aux tâches de conversation.

#### **4.4 Collecte des données**

Les données sont des conversations recueillies dans le cadre d'un plus grand projet de recherche qui vise la compréhension des interactions sociales entre des jeunes adultes québécois avec ou sans trouble développemental du langage (TDL) dans les cinq contextes présentés précédemment. Selon un entretien semi-directif, les partenaires de conversation interagissent dans deux tâches de conversation d'environ cinq minutes dans chacun des contextes d'interaction. L'objectif de cet entretien semi-directif est de s'assurer que l'interaction génère des changements de tour de parole.

La première tâche de conversation est inspirée d'une tâche initialement prévue pour les personnes aphasiques et leur partenaire de communication (Croteau et al., 2018). Les partenaires de communication doivent collaborer afin d'arriver à un consensus quant à la meilleure réaction à avoir face à une courte mise en situation. La seconde tâche de conversation est une discussion spontanée sur un thème choisi par l'équipe de recherche (inspiré des travaux de McFarland (2001), par exemple : « Dans quel type de quartier aimeriez-vous vivre plus tard? »). Un dilemme et un thème de conversation différents sont proposés pour chaque contexte et ces paires de tâches sont administrées aléatoirement dans les contextes selon les dyades. L'ordre de présentation des tâches de conversation alterne d'une dyade à l'autre : certaines dyades commencent avec la tâche des dilemmes alors que d'autres commencent avec la tâche de discussions spontanées.

Comme les tâches conversationnelles relèvent de capacités pragmatiques différentes, seule la tâche de conversation spontanée est analysée pour la présente étude. Pour cette tâche, il est attendu que les personnes soient en accord ou en désaccord sans pour autant argumenter l'une avec l'autre<sup>7</sup>. La discussion spontanée sur un sujet quelconque est une tâche conversationnelle

---

<sup>7</sup> Les items précis qui ont été utilisés pour cette tâche se trouvent à l'Annexe A.

typique dans la recherche sur la conversation (par exemple Kogure, 2007; Koudenburg et al., 2013; McFarland, 2001).

## **4.5 Méthodes d'analyses**

Cette étude vise à analyser des conversations dans les contextes en vidéoconférence et en réalité virtuelle entre les dyades sans TDL. Pour répondre à la question de recherche – à savoir « Comment les contextes d'interaction en vidéoconférence et en réalité virtuelle influencent la coordination entre les partenaires de communication ? » – et pour atteindre les deux objectifs fixés, nous utilisons un devis mixte, soit d'abord une analyse quantitative des silences et des chevauchements puis une analyse qualitative des moments problématiques dans la conversation qui contreviennent au principe « sans silence ni chevauchement ». La première analyse permet vise à atteindre le premier objectif, soit de comparer les marqueurs de la coordination dans chacun des contextes d'interaction, alors que la deuxième analyse vise à atteindre le deuxième objectif, soit de relever les caractéristiques propres à chacun des contextes d'interaction lorsque la conversation n'est pas fluide.

### **4.5.1 Préparation des données**

Tout d'abord, un nettoyage des fichiers vidéo pour les contextes d'interaction et les tâches ciblés s'impose. Les fichiers vidéo sont segmentés afin de ne conserver que les segments pertinents pour les analyses. Ainsi, les segments conservés débutent lorsque l'expérimentatrice pose la question-thème de la conversation aux partenaires de conversation et se terminent après le dernier énoncé d'un des locuteurs, soit avant que l'expérimentatrice reprenne la parole pour mettre fin à la discussion. Cette segmentation des fichiers se fait à l'aide du logiciel imovie qui permet d'enregistrer les fichiers en format vidéo (.mp4) et en format audio (.wav). Les fichiers audios en format .wav sont nécessaires pour l'analyse avec le logiciel ELAN (Wittenburg et al., 2006).

Une fois que les fichiers vidéo sont nettoyés et que les formats audios sont créés, ces fichiers doivent être annotés méthodiquement. Ces annotations se font à l'aide du logiciel ELAN (Wittenburg et al., 2006), souvent utilisé en analyse conversationnelle (Hepburn & Potter, 2021;

notamment dans Holler et al., 2018; Stivers et al., 2009). Les annotations concernent les segments de parole de chaque personne, les silences, les chevauchements et les backchannels.

#### 4.5.1.1 Les silences

Les silences sont définis comme une période durant laquelle il n’y a aucun signal sonore. La segmentation de la bande audio pour repérer les silences se fait de manière semi-automatique en deux étapes. La première étape de l’annotation des silences se fait à l’aide du *Silence Recognizer MPI-PL* du logiciel ELAN (Wittenburg et al., 2006). Le reconnaiseur de silences fait une segmentation automatique de la bande audio (à l’aide du fichier en format .wav) en distinguant les moments de silences des segments de parole ou de bruits. Les paramètres du reconnaiseur de silences sont ajustés de sorte qu’il crée des annotations pour les segments silencieux ayant une durée minimale de 50 ms et les segments non-silencieux d’au moins 10 ms. Puisque le reconnaiseur de silences n’est pas infaillible et que les fichiers audios peuvent contenir des impuretés (par exemple, le bruit d’une chaise qui bouge), la deuxième étape consiste à faire la vérification et l’ajustement des segments de silences repérés par le reconnaiseur de silences afin de s’assurer que ce soient réellement des silences.

#### 4.5.1.2 Les chevauchements

Les chevauchements sont définis comme tout moment où les deux personnes parlent en même temps (Schegloff, 2001). L’annotation des chevauchements utilise les segments non-silencieux annoté grâce à la segmentation semi-automatique décrite précédemment (avec le *Silence Recognizer MPI-PL* et une correction manuelle au besoin). Parmi ces segments non-silencieux, les segments qui ne sont pas des paroles sont repérés (par exemple, le bruit d’une personne qui tousse) pour ne conserver que des segments de parole. Ces segments de parole doivent ensuite être segmentés à nouveau en fonction des deux situations possibles : soit une seule personne parle soit les deux personnes parlent.

### 4.5.2 Analyses quantitatives

À partir des annotations, les deux indicateurs de la coordination (silences et chevauchements) sont comparés pour les contextes en présence, en vidéoconférence et en réalité virtuelle en termes de fréquence et de durée. Ces comparaisons permettent de mieux concevoir comment

les contextes influencent les interactions en déterminant s'il y a plus ou moins de chevauchements et/ou de silences, et si ces derniers sont plus ou moins longs dans un contexte ou dans l'autre. Pour vérifier la durée moyenne des silences des dyades en fonction des différents contextes d'interaction, nous utilisons une analyse de la variance (ANOVA) à mesures répétées à un facteur en utilisant la méthode d'inférence bayésienne à l'aide du logiciel JASP (JASP Team, 2023).

L'inférence bayésienne est une démarche statistique qui nous permet d'obtenir une mesure de l'évidence en faveur de l'hypothèse alternative et de l'hypothèse nulle. Comparativement à l'analyse fréquentielle (classique) selon laquelle la valeur de  $p$  mène au rejet ou au non-rejet de l'hypothèse nulle, l'inférence bayésienne indique plutôt les probabilités en faveur de chacune des hypothèses. L'inférence bayésienne offre une interprétation plus riche des données lorsqu'il y a un non-rejet de l'hypothèse nulle puisqu'elle permet de distinguer entre *l'absence d'évidence* et *l'évidence d'absence* (van den Bergh et al., 2020). Si le facteur de Bayes (BF) est proche de 1, alors les données observées montrent une probabilité équivalente pour l'hypothèse nulle et pour l'hypothèse alternative (il y a un manque d'évidence qui empêche de conclure en faveur d'un modèle ou d'un autre). À l'inverse, plus BF s'éloigne de 1, plus les données observées appuient une hypothèse plus qu'une autre : un  $BF_{10} > 1$  appuie l'hypothèse alternative, alors qu'un  $BF_{01} > 1$  appuie l'hypothèse nulle. Par conséquent, plus  $BF_{01}$  s'éloigne de 1 (en valeur positive), plus les données montrent une évidence en faveur de l'hypothèse nulle, soit l'évidence d'une réelle absence (van den Bergh et al., 2020).

L'approche bayésienne est pertinente pour répondre à notre question de recherche puisqu'elle permet de déterminer, dans le cas où les contextes ne sont pas différents, s'ils sont pour autant réellement similaires. Cette mesure est particulièrement pertinente pour la comparaison entre les contextes à distance et le contexte en présence : un BF fortement en faveur de l'hypothèse nulle nous permettrait de conclure qu'il existe une forte probabilité que ce contexte soit réellement similaire au contexte en présence. En comparaison, l'approche classique (les statistiques fréquentielles) mènerait uniquement au non-rejet de l'hypothèse nulle, soit à la conclusion que les données observées ne permettent pas de montrer que les contextes sont significativement différents ni pour autant similaires.

L'ANOVA à mesures répétées compare la durée moyenne des silences entre les contextes d'interactions et ce, pour chacune des dyades. Ainsi, l'analyse tient compte du fait que ce sont les mêmes dyades qui sont observées dans chaque contexte (soit le fait que les groupes sont dépendants). Conceptuellement, l'ANOVA à mesures répétées prend en considération les différences inter-individuelles, soit le fait que certaines dyades puissent avoir des conversations moins fluides (avec plus de silences ou de chevauchements) que d'autres simplement en raison des traits personnels des partenaires de conversation.

Pour utiliser l'ANOVA à mesures répétées, il faut d'abord s'assurer que les données respectent les postulats de l'analyse qui sont les mêmes en inférence bayésienne qu'en statistique classique. Ainsi, pour chaque analyse statistique rapportée dans la section des résultats, les indices permettant de vérifier la sphéricité (le test de Mauchly doit être non-significatif), la normalité de la distribution des résiduels (en observant le *G-G plot*) et la normalité de la distribution des données (le test de Shapiro-Wilk doit être non-significatif) sont vérifiés.

L'interprétation des résultats des analyses bayésiennes se déroule en deux temps. Dans un premier temps, il faut interpréter le facteur de Bayes (BF). Il existe différentes recommandations pour l'interprétation de la valeur du BF, nous nous basons sur celles de Kass et Raftery (1995) :

- un BF entre 1 et 3 représente une faible évidence (voir peu importante) ;
- un BF entre 3 et 20 représente la présence d'une évidence ;
- un BF entre 20 et 50 représente une forte évidence et ;
- un BF supérieur à 50 représente une évidence convaincante (très forte).

Dans un deuxième temps, si  $BF > 3$ , alors les comparaisons à postériori sont interprétées. Toutes les probabilités à postériori sont ajustées en fonction d'une probabilité a priori de l'hypothèse nulle fixée à 0,50 pour chaque comparaison ce qui permet de corriger pour l'accumulation de l'erreur due aux comparaisons multiples (Westfall et al., 1997).<sup>8</sup>

---

<sup>8</sup> Consulter van den Bergh et al. (2020) pour une synthèse récente de l'approche utilisée par le logiciel JASP pour la correction apportée pour les comparaisons multiples (à postériori).



### 4.5.3 Analyses qualitatives

L'analyse qualitative de ce mémoire combine l'analyse conversationnelle et l'analyse de contenu. Tel que mentionné dans le contexte théorique (section 2.1), l'analyse conversationnelle est l'une des méthodes typiquement utilisées pour étudier les interactions sociales et plus précisément les tours de parole (Hepburn & Potter, 2021). En ce sens, elle est la méthode la plus appropriée pour cette recherche. Cependant, les transcriptions très détaillées de l'analyse conversationnelle ne semblent pas nécessaires à la présente étude puisqu'elle ne comprend pas l'analyse de la prosodie.<sup>9</sup> De plus, l'objectif de notre analyse qualitative est moins de décrire dans les plus fins détails les interactions que d'arriver à comparer les événements problématiques entre les contextes d'interaction. En ce sens, l'analyse de contenu, avec la catégorisation hiérarchique des observations, nous permet de mieux répondre à l'objectif grâce à l'élaboration de thèmes et de catégories basés sur les observations qui peuvent ensuite être comparés d'un contexte à l'autre (Robert & Bouillaguet, 2007). Cela dit, l'interprétation des données observées pour créer les catégories plus larges se base sur les principes et concepts établis en analyse conversationnelle. En résumé, l'analyse conversationnelle permet d'interpréter les éléments de l'interaction alors que l'analyse de contenu permet d'organiser ces éléments en thèmes et en catégories utiles pour la comparaison des contextes d'interaction.

Avant d'entamer le processus de catégorisation de l'analyse de contenu, il faut suivre certaines étapes générales de l'analyse qualitative (cf. Aubin-Auger et al., 2008). D'abord, il faut s'assurer que la quantité de cas décrits est suffisante pour atteindre la saturation des données (soit le moment où les descriptions récoltées semblent suffisantes pour couvrir l'ensemble des phénomènes à l'étude). Ensuite, il faut choisir une méthode de codage. D'un côté, tel que présenté dans le contexte théorique, les silences et les chevauchements ont déjà été étudiés largement dans la littérature scientifique (entre autres dans Chowdhury et al., 2017; Heldner & Edlund, 2010; Lestary et al., 2018; Levinson & Torreira, 2015). Donc, un codage ouvert – qui serait uniquement basé sur le matériel analysé, sans a priori – ne serait pas justifié. D'un autre côté,

---

<sup>9</sup> Nous reconnaissons l'importance des indices prosodiques pour la coordination des tours de parole (voir par exemple dans Bögels & Torreira, 2021). Cependant, pour les besoins du présent mémoire, nous devons limiter nos observations à seulement certains éléments, donc les indices prosodiques ne sont pas analysés.

puisque le contexte en VR est récent, un codage strictement fermé où les catégories sont totalement prédéterminées à partir de la littérature serait tout aussi inapproprié. Par conséquent, le codage mixte est favorisé puisqu'il se base sur la littérature existante tout en permettant le retrait et/ou l'ajout de certaines catégories (ou thèmes) au besoin. Comme la VR est un contexte d'interaction émergent, tout l'intérêt d'utiliser une méthode qualitative repose sur cette souplesse qui nous permet d'avoir une approche exploratoire rigoureuse.

Pour procéder à l'analyse de contenu, il faut d'abord décrire de la façon la plus systématique et objective possible les éléments ciblés (Robert & Bouillaguet, 2007). Dans le cadre de cette étude, nous limitons nos observations aux énoncés verbaux ainsi qu'à certains indices non-verbaux perceptibles – soit les mouvements de la tête, des mains, des bras, des épaules et du corps entier ainsi que les expressions faciales. En raison des limites des enregistrements vidéo, la direction du regard n'est pas analysée avec précision.<sup>10</sup> Une fois les descriptions complétées, un travail d'interprétation des données permet de rassembler les éléments décrits sur la base de leurs similarités dans des catégories, puis dans des thèmes plus larges. Ainsi, l'analyse de contenu permet de classer les observations dans des catégories qui peuvent ensuite être utilisées pour comparer les contextes d'interaction entre eux. En effet, une fois les thèmes et catégories établis, les contextes d'interaction peuvent être comparés les uns aux autres en comptant le nombre de situations qui s'appliquent à une certaine catégorie. Par exemple, si dans la catégorie « résolution du chevauchement par un silence », on compte moins de cas qui se produisent en VR qu'en vidéoconférence, alors on pourrait conclure que les contextes d'interaction sont différents en ce qui a trait à l'utilisation du silence comme stratégie de résolution de chevauchements.<sup>11</sup> Par

---

<sup>10</sup> Étant donné que l'objectif principal de cette étude ne concerne pas directement les regards, le matériel utilisé pour l'enregistrement vidéo n'est pas spécialisé pour en faire la détection précise. Donc, il est parfois difficile (surtout dans les interactions en présence) de détecter systématiquement la direction du regard des personnes participantes. C'est pourquoi nous avons choisi de ne pas analyser les regards de façon détaillée. Cela dit, comme mentionné dans le contexte théorique, les regards sont des indices importants pour la coordination de la conversation. Pour cette raison, des remarques à ce sujet sont tout de même intégrées dans la discussion en prenant en considération l'observation générale que, du moins pour notre étude, le contexte en vidéoconférence permet une certaine transmission du regard que la VR ne permet pas.

<sup>11</sup> Il est à noter que cet exemple n'est pas nécessairement représentatif des données. Le nom de la catégorie de même que la comparaison et la conclusion servent uniquement à illustrer concrètement l'intérêt d'utiliser l'analyse de contenu pour répondre à l'objectif. Les catégories réelles ainsi que les comparaisons sont présentées dans la section des résultats (5.3).

conséquent, l'analyse de contenu permet d'étudier les interactions sociales (plus précisément les événements problématiques) en mettant en valeur des éléments de comparaison entre les contextes. Cela permet de répondre au deuxième objectif de la recherche qui vise à comparer les contextes d'interactions lorsqu'un bris dans la fluidité de la conversation survient.

#### 4.5.3.1 Événements problématiques ciblés par les analyses qualitatives

Comme notre étude porte sur la coordination de la conversation, les événements problématiques sont ceux qui cause un bris dans la fluidité des échanges : ce sont des chevauchements et des silences qui violent le principe de conversation « sans silence ni chevauchement ». Il y a une analyse qualitative pour les longs silences et une autre pour les chevauchements.

À partir des données quantitatives, ce sont les silences qui sont très longs qui sont choisis pour l'analyse qualitative. Selon l'observation des quartiles dans chacun des contextes d'interaction, il semble que les silences les plus longs sont ceux qui sont les plus influencés par les contextes d'interaction puisque la différence entre les quartiles augmente avec l'augmentation des quartiles. Plus précisément, le Tableau 2, montre que les premiers quartiles égalent 220 ms, 210 ms et 239 ms alors que les troisièmes quartiles sont plutôt à 675 ms, 625 ms et 905 ms pour le contexte en présence, en vidéoconférence et en VR respectivement. Comme la différence entre les troisièmes quartiles (variation de 230 ms) est plus grande que celle entre les premiers quartiles (variation de 29 ms), nous avons choisi d'explorer qualitativement les longs silences. Le seuil minimal pour les longs silences est établi à 1 165 ms, ce qui correspond au 90<sup>e</sup> centile (soit les silences qui sont plus longs qu'au moins 90 % des autres silences). De plus, ce seuil minimal pour les longs silences fait écho aux travaux antérieurs montrant que les silences de plus d'une seconde sont ceux qui sont problématiques et qui perturbent le flux de la conversation (Jefferson, 1989; Koudenburg et al., 2013).

Les chevauchements sont des sources de bris dans la communication qui doivent généralement être réparés (à l'exception des BC tel que mentionné dans le contexte théorique). Afin d'approfondir la manière dont les contextes d'interaction influencent la coordination liée aux chevauchements, l'analyse qualitative des chevauchements décrit leurs différentes

caractéristiques en incluant les contextes dans lesquels ils apparaissent, le type de chevauchement ainsi que leur dénouement (Hayashi et al., 2013; Schegloff et al., 1977).

L'interaction avant, pendant et après chaque événement problématique (les longs silences et les chevauchements) est décrite en détails en se basant sur les enregistrements dans l'environnement virtuel (vidéo ou VR), mais aussi dans l'environnement physique réel. L'observation de l'environnement réel permet de rendre compte des erreurs dues aux limites de la technologie (par exemple les mauvaises correspondances entre les mouvements réels d'un individu et ceux de son avatar).

Ces analyses permettent d'atteindre les objectifs du mémoire soit de présenter les différences quantitatives entre les différents indices de la coordination et de mettre en lumière les caractéristiques propres à chacun des contextes qui sont en jeu lorsqu'un problème de fluidité survient.

## Chapitre 5 – Résultats

### 5.1 Échantillon

L'échantillon est composé de 16 adultes âgés entre 20 et 30 ans ayant participé aux conversations en paires (huit dyades). Le Tableau 1 présente une synthèse des données descriptives en lien avec chacune des conversations.

**Tableau 1**

*Données descriptives des conversations en fonction des dyades*

Conversations	Âge (ans)	Sexe	Durée (mm : ss,ms)		
			Présence	Vidéo-conférence	Réalité virtuelle
ID-01	26	F	04 : 45,05	03 : 12,26	02 : 25,59
	26	F			
ID-02	23	F	04 : 12,79	05 : 17,12	02 : 23,32
	23	F			
ID-03	24	F	04 : 10,80	03 : 39,10	04 : 41,16
	24	M			
ID-04	29	F	06 : 06,87	08 : 19,39	06 : 25,31
	28	M			
ID-05	25	F	05 : 01,17	05 : 34,88	04 : 48,36
	26	F			
ID-06	20	F	05 : 17,72	05 : 50,07	04 : 22,01
	21	F			
PI-01	23	M	04 : 52,76	05 : 51,71	04 : 52,76
	30	M			
PI-02	25	M	03 : 22,48	03 : 17,99	02 : 29,41
	20	M			
Moyenne	25		04 : 39,86	05 : 07,81	04 : 18,34
<b>Total</b>		<b>F : 10</b> <b>M : 6</b>	<b>37 : 18,89</b>	<b>41 : 02,52</b>	<b>34 : 26,75</b>

Dix personnes s'identifient au sexe féminin (six s'identifient au sexe masculin). La majorité sont aux études supérieures (N=11) alors que les autres sont sur le marché du travail (N=4) (une personne est à la fois étudiante et travailleuse). Un seul individu mentionne avoir utilisé souvent

la réalité virtuelle avec un casque d'immersion alors que tous les autres indiquent avoir utilisé cette technologie « peu de fois » (N=5) ou « jamais » (N=10). Toutes les personnes indiquent avoir déjà utilisé la vidéoconférence au moins « peu de fois ». Nous avons analysé huit discussions d'environ cinq minutes dans chacun des contextes d'interaction, pour un total de 113 minutes d'enregistrement, pour lesquelles 2 390 silences et 6 881 segments de parole ont été annotés.

## **5.2 Les contextes d'interaction et les indices de la coordination**

### **5.2.1 Les silences**

#### 5.2.1.1 La fréquence des silences

Pour l'ensemble des contextes annoté, 2 389 silences sont observés. 1 583 de ces silences sont des pauses (c'est-à-dire que le silence survient à l'intérieur du tour de parole de la même personne) et 806 sont des intervalles (c'est-à-dire qu'un nouveau locuteur prend la parole après le silence). Les données descriptives des silences pour chaque contexte en fonction des dyades sont présentées au Tableau 2. Le nombre de silences pour une minute (N/min) permet de comparer la fréquence des silences en prenant en considération que les conversations n'ont pas exactement la même durée dans chacun des contextes d'interaction. Ces données montrent que les conversations en vidéoconférence (20,49 N/min), en réalité virtuelle (21,22 N/min) et en présence (21,89 N/min) ont un nombre de silences qui est semblable. De plus, tous les contextes présentent des proportions similaires de pauses et d'intervalles, soit davantage de pauses que d'intervalles.

**Tableau 2***Distribution et durée de tous les silences*

Contextes	Fréquences			Moyenne (ms)	Mode (ms)	Quartiles (ms)		
	N	N/min	N/N <sub>total</sub> (%)			1 <sup>er</sup>	2 <sup>e</sup>	3 <sup>e</sup>
<b>Présence</b>	<b>817</b>	<b>21,89</b>	<b>34</b>	<b>511,67</b>	<b>110</b>	<b>220</b>	<b>370</b>	<b>675</b>
Pauses	502	13,45		546,42				
Intervalles	315	8,44		455,98				
<b>Vidéo- conférence</b>	<b>841</b>	<b>20,49</b>	<b>35</b>	<b>501,13</b>	<b>140</b>	<b>210</b>	<b>377</b>	<b>625</b>
Pauses	593	14,45		472,11				
Intervalles	248	6,04		570,52				
<b>Réalité virtuelle</b>	<b>731</b>	<b>21,22</b>	<b>31</b>	<b>648,97</b>	<b>260</b>	<b>239</b>	<b>540</b>	<b>905</b>
Pauses	488	14,17		552,29				
Intervalles	243	7,05		843,12				
<b>Total</b>	<b>2 389</b>	<b>21,18</b>	<b>100</b>	<b>549,90</b>	<b>140</b>	<b>220</b>	<b>400</b>	<b>740</b>
Pauses	1 583	14,03		520,45				
Intervalles	806	7,15		607,94				

Le Tableau 3 présente les données descriptives associées aux longs silences. Comme mentionné dans le plan d'analyse à la section précédente (section 4.5.3), les longs silences correspondent aux silences ayant une durée de plus 1 165 ms ce qui correspond au 90<sup>e</sup> centile. Ces longs silences représentent 10 % de tous les silences (N=240), dont 130 sont de longues pauses et 110 sont de longs intervalles. Les longs silences semblent un peu plus présents dans les conversations en réalité virtuelle (3,22 N/min) que dans les conversations en vidéoconférence (1,54 N/min) ou en présence (1,77 N/min). Les longs silences sont issus davantage des conversations en réalité virtuelle (46 %) que des autres contextes d'interaction (28 % des longs silences sont issus des conversations en présence et 26 % se trouvent dans celles en vidéoconférence).

**Tableau 3***Distribution et durée des longs silences*

Contextes	Fréquences			Moyenne (ms)
	N	N/min	N/N <sub>total</sub> (%)	
<b>Présence</b>	<b>66</b>	<b>1,77</b>	<b>28</b>	<b>1 584,89</b>
Pauses	44	1,18		1 573,73
Intervalles	22	0,59		1 607,23
<b>Vidéo-conférence</b>	<b>63</b>	<b>1,54</b>	<b>26</b>	<b>1 637,30</b>
Pauses	36	0,88		1 655,00
Intervalles	27	0,66		1 613,70
<b>Réalité virtuelle</b>	<b>111</b>	<b>3,22</b>	<b>46</b>	<b>1 633,42</b>
Pauses	50	1,45		1 671,54
Intervalles	61	1,77		1 602,18
<b>Total</b>	<b>240</b>	<b>2,13</b>	<b>100</b>	<b>1 621,10</b>
Pauses	130	1,15		1 633,85
Intervalles	110	0,98		1 606,02

#### 5.2.1.2 La durée des silences

Tous les silences observés varient entre 50<sup>12</sup> et 5 700 ms, et la durée moyenne (M) de ces silences est de 549,90 ms. L'ANOVA bayésienne à mesures répétées pour l'ensemble des silences montre que les différents contextes d'interaction ont effectivement un effet sur la durée des silences ; la durée moyenne des silences est différente d'un contexte à l'autre ( $BF_{10} = 19,91$ ). Plus précisément, les analyses à posteriori montrent une tendance vers une faible évidence d'une différence entre le contexte en réalité virtuelle et les contextes en vidéoconférence (probabilité corrigée de 2,99) et en présence (probabilité corrigée de 2,96). Tel qu'attendu, les silences en VR tendent à être plus longs (M = 648,97 ms ; Mode = 260 ms ; Médiane = 540 ms) que ceux en

<sup>12</sup> Rappelons que 50 ms est la valeur minimale qui a été utilisée pour la détection semi-automatique des silences. Ce seuil de précision semble adéquat pour notre étude et coïncide avec les données trouvées dans la littérature (Chowdhury et al., 2017).



vidéoconférence (M = 500,89 ms ; Mode = 140 ms ; Médiane = 376 ms) et en face-à-face (M = 511,67 ms ; Mode = 110 ms ; Médiane = 370 ms). Cela dit, cette analyse ne permet pas de dire qu'il y a une différence entre les silences en vidéoconférence et en présence (probabilité corrigée de 0,21).

Comme mentionné plus tôt, l'ANOVA bayésienne à mesures répétées permet d'étudier plus en profondeur le « manque de différence » (ou les situations où les statistiques classiques se contenteraient du « non-rejet de l'hypothèse nulle »). Dans le cas présent, on cherche à savoir s'il y a réellement une similarité entre les contextes en présence et en vidéoconférence ou bien si nos observations ne nous permettent simplement pas d'arriver à une conclusion convaincante. La valeur de  $BF_{01} = 20,03$  montre une forte évidence en faveur de l'hypothèse nulle. De plus, la probabilité corrigée de 4,83 pour la comparaison à posteriori des contextes en vidéoconférence et en présence montre une évidence faible en faveur de la similarité de ces contextes.

### **5.2.2 Les chevauchements**

630 chevauchements ont été observés. La majorité de ces chevauchements, soit 53,33 % sont des BC (N = 337). Comme il est indiqué dans la section précédente, les BC seront exclus des données afin de concentrer les observations sur les chevauchements, qui sont davantage associés à un problème dans la conversation. Donc, le Tableau 4, présente la distribution des chevauchements qui ne sont pas des BC : 293 chevauchements, dont 200 chevauchements inter-locuteur (chevauchements suivis d'un changement de locuteur) et 93 chevauchements intra-locuteurs (chevauchements qui ne sont pas suivis d'un changement de locuteur).

**Tableau 4***Distribution des chevauchements selon les contextes d'interaction*

Contextes	Chevauchements			<i>Backchannel</i>
	Inter-locuteur	Intra-locuteur	Total	
Présence	85 (2,28)	41 (1,10)	126 (3,38)	136 (3,64)
Vidéoconférence	51 (1,24)	21 (0,51)	72 (1,75)	110 (2,68)
Réalité virtuelle	64 (1,86)	31 (0,90)	95 (2,76)	91 (2,64)
Total	200 (1,77)	93 (0,82)	293 (2,60)	337 (2,99)

*Note.* Les valeurs présentées entre parenthèses sont des ratios de l'occurrence (N) de chaque événement pour une minute. Cela permet de prendre en considération le fait que les interactions dans les différents contextes n'ont pas exactement la même durée. Les BC sont présentés à titre informatif.

Il y a plus de chevauchements qui se produisent dans le contexte en présence (3,38 N/min) comparativement aux contextes en vidéoconférence (1,75 N/min) et en réalité virtuelle (2,76 N/min). Comme pour les silences, l'ANOVA bayésienne à mesures répétées est utilisée pour vérifier la valeur de ces observations. Les données ne permettent pas d'obtenir une forte évidence en faveur de l'hypothèse alternative ( $BF_{10} = 2,60$ ) ni en faveur de l'hypothèse nulle ( $BF_{01} = 0,39$ ). On observe une tendance en faveur d'une différence entre les contextes (hypothèse alternative), mais nos observations semblent insuffisantes pour arriver à une conclusion robuste puisque les BF sont trop proches de 1.

### 5.2.3 Conclusion sur le premier objectif

Pour répondre au premier objectif de recherche, les silences en réalité virtuelle sont effectivement plus longs que ceux en vidéoconférence. Ces données nous indiquent plus globalement que la réalité virtuelle influence différemment la coordination entre les partenaires de conversation que la vidéoconférence. Par ailleurs, l'inférence bayésienne appuie l'hypothèse en faveur d'une similarité entre la vidéoconférence et la conversation en face-à-face, et celle en faveur d'une différence entre la réalité virtuelle et la conversation en présence. Comparativement à la réalité virtuelle, la vidéoconférence serait donc le contexte qui ressemble le plus au contexte naturel en présence en ce qui a trait aux silences. Pour ce qui est des chevauchements, ceux-ci

semblent plus nombreux en présence et en VR que dans les conversations en vidéoconférence. Cependant, un nombre supplémentaire d'observations est nécessaire avant de pouvoir émettre une conclusion convaincante quant à la fréquence des chevauchements dans les différents contextes d'interaction.

### **5.3 Les contextes d'interaction et les événements problématiques**

Pour comprendre davantage l'influence des contextes d'interaction sur la coordination entre les interlocuteurs, la présente section porte sur les résultats de l'exploration qualitative des événements qui comportent une violation du principe « sans silence ni chevauchement ». Plus précisément, cette analyse qualitative vise à mettre en lumière les caractéristiques associées aux événements qui semblent particulièrement problématiques (les longs silences et les chevauchements) ainsi que la manière dont les partenaires de conversation arrivent à reprendre le rythme de leur conversation à la suite de ces événements.

#### **5.3.1 Les silences**

Tel que mentionné dans la section décrivant le plan des analyses qualitatives, seuls les silences de plus de 1 165 ms (correspondant au 90<sup>e</sup> centile) ont été décrits qualitativement, soit un total de 240 silences ayant des durées qui varient entre 1 165 et 5 700 ms. Le Tableau 5 présente une synthèse des différents thèmes et de leurs catégories qui décrivent les longs silences.

**Tableau 5***Synthèse des thèmes et des catégories des longs silences*

Thèmes	Catégories	Descriptions
Les causes apparentes principales	Réflexion	Le silence semble être utilisé par les partenaires pour réfléchir à leur énoncé (recherche d'un mot ou d'une idée).
	Absence de prise de parole	Après un TRP, personne ne prend la parole sans raison apparente.
	Résolution d'un bris de communication	Le silence semblent être utilisé pour rétablir le rythme de la conversation après un bris.
L'interaction pendant le silence	Le maintien de l'interaction	Pendant le silence, les partenaires démontrent clairement leur engagement continu dans l'interaction par leur gestuelle.
	L'absence d'interaction	Pendant le silence, il n'y a pas d'échange non-verbal entre les partenaires.

## 5.3.1.1 Causes apparentes des longs silences

Le premier thème s'intéresse aux causes apparentes des long silences dans chacun des contextes d'interaction. Ce thème est subdivisé en trois catégories : les réflexions, les bris de communication et l'absence de prise de parole.

La première caractéristique, la plus récurrente dans l'observation des longs silences, est qu'ils sont souvent la conséquence d'une réflexion (N=138 ; 58 %). Ce sont des silences qui sous-entendent que le locuteur et/ou l'interactante cherche ses mots ou ses idées. Ces silences peuvent se produire après un TRP lorsque les partenaires semblent avoir épuisé le sujet de conversation ou après une hésitation dans le discours du locuteur. L'hésitation est reconnaissable par l'interruption brusque d'un énoncé, par la présence d'une particule de remplissage comme « hum » et « heu » à la fin de l'énoncé ou encore par l'allongement de la dernière syllabe. Les silences de réflexion sont aussi présents à la suite d'un énoncé qui appelle une réponse de l'interactante, lorsque le locuteur attribue le tour de parole à l'interactante en posant une

question directe à celle-ci. Dans ce cas, le silence est issu de la réflexion de l'interactante et non du locuteur. Par ailleurs, les silences associés à une réflexion s'accompagnent généralement d'un ou de plusieurs éléments de communication non-verbale tels que le retrait du regard pour fixer le plafond ou le sol ou encore l'arrêt de la gesticulation du locuteur. Dans ce cas, le locuteur reste figé avec une gesticulation non-complétée durant le silence et reprend (ou termine) sa gesticulation généralement au moment où il reprend la parole (si le silence est une pause) ou il abandonne sa gesticulation lorsque l'autre prend la parole (si le silence est un intervalle). Les silences de réflexion correspondent à 70 % des longs silences en présence et à 71 % de ceux en vidéoconférence alors qu'ils ne représentent que 42 % des longs silences en VR. L'Extrait 1 présente un exemple typique des longs silences qui semblent être la conséquence d'une réflexion.

### **Extrait 1**

#### *Exemple de silence de réflexion*

01 A : Hum,  
02 (1,62)  
03 A : Peut-être un mix des deux ?  
04 (1.58)  
05 B : [ O k ] ((hoche la tête))  
06 A : [tsais] comm::::e ((lève les yeux au plafond))

Dans cet exemple, le locuteur A prend de longues pauses (aux lignes 02 et 04) pour réfléchir à sa réponse. Les signes de réflexion sont la présence de la particule « hum » (à la ligne 01), l'hésitation au niveau de la prosodie (à la ligne 03) et le retrait du regard pour fixer le plafond (à la ligne 06).

La deuxième catégorie regroupe les silences pour lesquels il y a une absence de prise du tour de parole après la fin d'un énoncé. Pour ces silences, il n'y a pas de causes apparentes autre que le fait que personne ne réclame le tour de parole après la complétion du tour du locuteur. Donc, ces silences se produisent après un TRP et ils n'impliquent aucun signe de réflexion et aucun signe de désaccord entre les partenaires qui pourraient les justifier. Ces silences surviennent avec des proportions similaires dans chaque contexte, soit 24 % en présence, 24 % en vidéoconférence, 28 % en VR.

## Extrait 2

### *Exemple d'absence de prise de parole après un TRP*

01 C2 Elle est (0.37) au centre de tout pis c'est com- (.) c'est ça qu'tu  
02 veux quand t'es jeu:ne là surtout [qu'elle a-] elle conduit pas là  
03 [ouai:s]  
04 (1.46)  
05 C1 Ouais (.) tu veux avoir le transport en commun accessible pis comme  
tout tes besoins primaires rapidement comblé

L'Extrait 2 illustre plus clairement un exemple de long silence qui se produit après TRP. Le locuteur C2 arrive à la fin de son énoncé (ligne 02), mais C1 ne prend la parole (ligne 05) qu'après un long silence (ligne 04).

La troisième catégorie, les silences de résolution, rassemble les silences qui se produisent à la suite d'un bris de communication, plus précisément après un chevauchement. Les chevauchements seront davantage discutés dans la section suivante, mais il va sans dire que les silences sont fréquents et communs à la suite d'un bris de communication. Cela dit, les longs silences qui surviennent à la suite d'un chevauchement sont presque uniquement dans le contexte en VR. En effet, alors qu'il n'y a aucun silence de résolution dans les conversations en présence et un seul en vidéoconférence, il y en a 20 dans le contexte en VR. Cela représente 18 % des longs silences en VR.

Finalement, les autres longs silences qui ne sont pas classés parmi les plus grandes catégories présentées précédemment sont dus à des désaccords entre les partenaires (N=10; 4 %) ou à des distractions de la VR (N=9; 3,75 %). Les désaccords ne seront pas discutés davantage étant donné leur faible occurrence dans chacun des contextes d'interaction (quatre en présence, deux en vidéoconférence et quatre en VR). En effet, puisque les silences qui sont liés aux désaccords (ou à une réponse moins préférée) sont peu nombreux, il ne semble pas avisé de tenter de tirer quelque conclusion quant à leur présence dans les différents contextes. Dans le cas des distractions dues à la VR, ce sont des situations qui sont évidemment exclusives au contexte en VR. Ces silences impliquent que les partenaires font et/ou disent des choses qui sont liées

directement avec l'environnement virtuel, et c'est pourquoi il est légitime de croire que ces silences n'auraient pas eu lieu (du moins pas sous cette forme) dans les autres contextes. Dans la plupart des cas, les partenaires font un mime qui est reproduit par l'avatar, et ce qui est dit avant ou après le silence n'a aucun lien avec le cours de la conversation. À d'autres moments, les interlocuteurs semblent distraits par un problème lié à la correspondance des mouvements entre un avatar et la personne qui l'incarne. La Figure 1 est un exemple de silence qui est engendré par l'environnement virtuel qui cause une distraction.

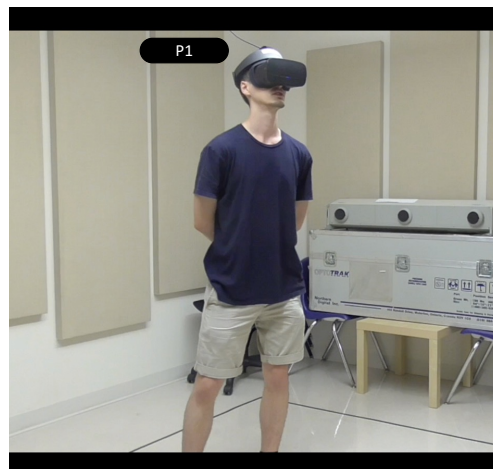
**Figure 1**

*Exemple d'un problème engendré dans la réalité virtuelle*



**Figure 2**

*Position réelle du participant P1*



Dans la Figure 1, le bras de P1 est placé dans une position étrange qui est due à un problème de concordance entre la VR et la position réelle des mains de la personne qui incarne l'avatar. En effet, comme montre la Figure 2, P1 se tient simplement debout avec les bras croisés dans son dos.

Il y a un long silence après la fin du tour de parole du locuteur P1. Après le silence, le locuteur P2 pointe P1 (comme sur la Figure 1) et demande à P1 ce qu'il fait avec son bras. Il est possible que P2 n'ait pas pris la parole immédiatement après la fin du tour de P1 puisqu'il était distrait par le problème qu'il percevait : il tentait de comprendre ce que l'autre faisait avec son corps.

Les données observées relèvent des différences et des similarités entre les contextes quant aux causes apparentes des longs silences. De façon générale, la VR semble être plus différente du contexte en présence que la vidéoconférence. Comparativement aux conversations en VR, celles en vidéoconférence (comme en présence) n'utilisent jamais (ou presque) les longs silences pour résoudre un chevauchement et elles sont évidemment sans distractions due à la VR.

#### 5.3.1.2 L'interaction pendant les silences

Durant les silences, les partenaires de conversation sont rarement immobiles. En effet, l'absence de parole n'implique pas nécessairement l'absence de communication. Pour cette raison, le deuxième thème porte sur la description des longs silences en fonction de ce qui se passe pendant ces silences. Il faut préciser que les gestes et les mouvements peuvent être différents d'un contexte à l'autre. D'un côté, la vidéoconférence permet la perception réelle des mouvements de la tête et des épaules, des expressions faciales ainsi que de certains gestes des mains (lorsque la gesticulation est près du visage). D'un autre côté, le système de VR permet la perception de certains mouvements transmis par l'entremise de l'avatar : des mouvements de la tête (comme hocher de haut en bas ou secouer de gauche à droite), des mouvements des mains (comme pointer et gesticuler) et du corps (comme avancer, reculer et sauter).

Les observations se concentrent sur les mouvements perceptibles de la tête, des mains et du corps et sont regroupés selon deux catégories : les mouvements qui visent le maintien de l'interaction et les autres mouvements.



La première catégorie se concentre sur les gestes des partenaires de communication durant les silences qui assurent le maintien de l'interaction. La catégorie regroupe tous les silences pour lesquels au moins un des deux partenaires effectue un comportement qui a une fonction communicationnelle ou un comportement qui indique qu'au moins un des deux partenaires est toujours engagé activement dans la conversation. Parmi les comportements ayant une fonction de communication précise, il y a des éléments qui sont perceptibles en vidéoconférence et en VR comme les différents mouvements de la tête (hocher de haut en bas ou secouer de gauche à droite) qui sont souvent signes de BC et la gesticulation des mains (gesticulation continue ou figée à mi-chemin) qui indique que le locuteur désire maintenir son tour de parole. Certains comportements sont perceptibles uniquement en vidéoconférence comme la présence d'une expression faciale (par exemple, la surprise ou le dégoût). Il y a aussi des situations de silence durant lesquelles une personne hausse les épaules (perceptible seulement en vidéoconférence et en présence) ou lève les mains dans les airs pour signifier qu'elle ne sait pas. Dans les comportements qui témoignent d'un engagement actif dans la conversation, on fait référence aux situations extraordinaires où les partenaires miment quelque chose, dansent ou encore se font une poignée de main. Ces comportements ne transmettent pas une information précise (comparativement au hochement de la tête qui est signe d'approbation), mais leur exécution souligne tout de même le maintien de l'interaction, et donc une certaine coordination entre les partenaires qui maintiennent leur engagement dans l'interaction sociale. En VR, il y a une moins grande proportion de longs silences durant lesquels une interaction non-verbale est clairement maintenue (N=40; 36,04 %) comparativement aux conversations en vidéoconférence (N=30; 47,62 %) et celles en présence (N=36; 54,55 %).

La deuxième catégorie regroupe les silences qui ne présentent pas de gestes qui assurent le maintien de l'interaction. Ce sont des silences durant lesquels il peut y avoir une absence de gestes ou encore la présence de gestes ou de mouvements qui ne contribuent pas au maintien de l'interaction. Il y a absence de gestes lorsque les partenaires sont simplement immobiles dans une position *neutre*. On entend par « neutre » que les mains et le corps des partenaires sont au repos (contrairement à des situations dans lesquelles les mains seraient figées à mi-chemin d'une gesticulation). Il y en a 63,96 % en VR comparativement à 52,38 % en vidéoconférence et 45,45 %

en présence. Les gestes qui ne contribuent pas à la communication de façon explicite sont des comportements qui sont surtout observés dans les conversations en réalité virtuelle. Ces gestes comprennent différents problèmes de correspondance entre l'avatar en VR et la personne qui l'incarne. Il est souvent question d'un bras qui se lève brusquement dans les airs dans une position étrange alors que la personne ne bouge pas dans la réalité (revoir l'exemple de la Figure 1 comparativement à la Figure 2). Parmi les gestes qui ne contribuent pas à l'interaction, on compte aussi les mouvements en VR qui semblent servir à l'exploration de l'environnement et non à l'échange d'information. Par exemple, une personne qui fait sauter son avatar sur place ou qui court dans toutes les directions durant le silence exécute des mouvements qui peuvent avoir un effet sur le déroulement de la conversation, mais ces comportements ne semblent pas présenter un objectif au niveau de la communication. Ces comportements semblent plutôt être des distractions autant pour la personne qui exécute le mouvement et que pour celle qui le perçoit. D'un côté, la personne qui effectue le comportement a une attention qui est divisée entre l'exécution du mouvement (l'exploration des possibilités en VR) et la conversation. D'un autre côté, la personne qui perçoit les mouvements peut être distraite par le fait qu'ils soient inconvencionnels et/ou inattendus et par le besoin de les interpréter (afin de comprendre s'ils sont produits pour une raison communicationnelle ou non). Cette distraction risque d'influencer le rythme de la conversation et nuire à l'échange rapide et efficace des tours de parole. Les Figures 3 et 4 comparent deux longs silences en VR. Dans la Figure 3, il y a clairement une interaction entre les partenaires K et J qui se font une poignée de mains alors que dans la Figure 4, B fait sauter son avatar sur place, mais ce mouvement ne semble pas avoir de but au niveau de la communication : il ne sert pas à maintenir l'interaction. En ce sens, ce comportement durant le long silence semble simplement dû au fait que B explore ce que le monde virtuel lui permet de faire.

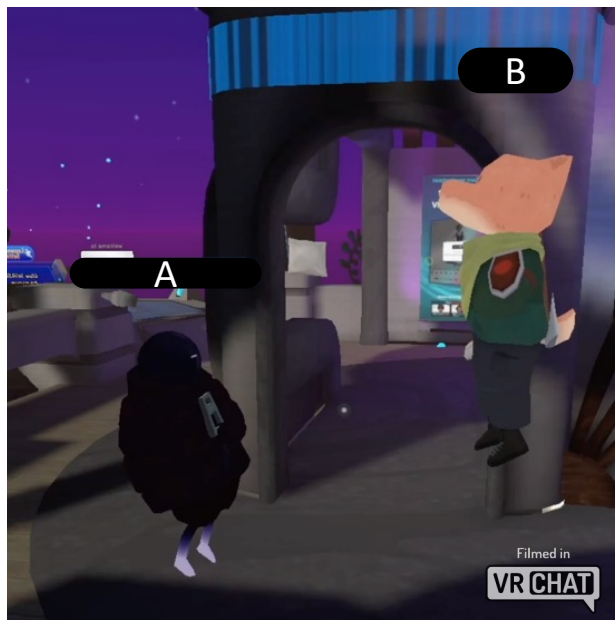
**Figure 3**

*Exemple de maintien de l'interaction pendant le long silence grâce aux gestes*



**Figure 4**

*Exemple de gestes qui ne maintiennent pas l'interaction durant le long silence*



### **5.3.2 Les chevauchements**

Nous avons analysé les 293 chevauchements qui ne sont pas engendrés par des BC pour chaque contexte d'interaction. Le Tableau 6 présente un récapitulatif des différents thèmes et de leurs catégories ainsi que la fréquence d'apparition de ces catégories pour les contextes en vidéoconférence, en présence et en réalité virtuelle.

La description des chevauchements est divisée en trois grands thèmes, soit le contexte dans lequel le chevauchement apparaît, le type de chevauchement et finalement le dénouement du chevauchement (sa résolution). Ces thèmes et les diverses catégories qu'ils contiennent sont issus de travaux antérieurs (notamment Levinson & Torreira, 2015) et des phénomènes observés dans nos données.

**Tableau 6***Synthèse des thèmes et des catégories de chevauchements*

Thèmes	Catégories	Descriptions
Contextes d'apparition	Après un silence	Le chevauchement se produit au maximum 200 ms après un silence.
	Près d'un TRP	Le chevauchement débute au maximum 500 ms avant ou après un TRP.
	Après une discontinuité	Le tour de parole qui est chevauché est discontinu (p.ex., énoncé saccadé, ralentissement du rythme).
	Au milieu du tour	Le chevauchement se produit lorsque l'interactante parle alors que le locuteur n'a pas terminé son tour et qu'il n'y a pas de raison de croire qu'un TRP approche.
Types de chevauchement	Départ simultané	Les deux partenaires prennent la parole en même temps.
	Reprise/ Ajout du locuteur	Le locuteur reprend la parole après un TRP alors que son partenaire a déjà commencé à parler.
	Commentaire sur la VR	Le chevauchement est exclusivement associé à un événement de l'environnement virtuel.
Dénouement du chevauchement	Abandon/ Interruption	Au moins un des deux partenaires arrêtent de parler.
	Silence	Les deux partenaires cessent de parler directement après le chevauchement ou peu de temps après.
	Répétition	Une ou plusieurs des syllabes (ou mots) prononcées lors du chevauchement sont répétées au début de l'énoncé suivant le chevauchement.

## 5.3.2.1 Le contexte d'apparition du chevauchement

Le premier thème, le contexte d'apparition, rassemble les caractéristiques qui permettent de mieux comprendre à quel moment, durant l'interaction, le chevauchement se produit. Ce thème

se subdivise en quatre catégories : (1) après un silence, (2) après un TRP, (3) après discontinuité dans le tour précédent et finalement (4) avec la fin du tour.

La première catégorie, « après un silence », comprend les chevauchements qui débutent au maximum 200 ms après un silence. Il y a moins de chevauchement qui se produisent après un silence dans le contexte en présence (15,87 %) que dans les contextes en vidéoconférence (35,62 %) et en VR (43,16 %).

La deuxième catégorie, « près d'un TRP », comprend les chevauchements qui débutent au maximum 500 ms avant ou après un TRP. Le chevauchement qui se produit près d'un TRP est généralement engendré par un ajout du locuteur ou plutôt la reprise du locuteur après la fin probable de son tour de parole. À l'inverse des chevauchements qui se produisent après un silence, ceux qui se produisent après un TRP sont davantage présents dans les conversations en présence (50,00 %) et dans les conversations en vidéoconférence (38,89 %) que dans les conversations en VR (17,89 %). L'Extrait 3 (aux lignes 03 et 09) illustre comment, en face-à-face, les partenaires semblent à l'aise d'échanger leur tour de parole en chevauchement avec la fin du tour du locuteur précédent.

### **Extrait 3**

#### *Exemple d'un échange en face-à-face*

01 K: Exactement (.) des chats ça se nettoie tout seul  
02 [moi mon chat je l'ai jamais nettoyé]  
03 J: [mais ça c't'un bon poi:nt] si c'est auto-nettoyant t'as  
04 pas besoin d'l'amener au n- au nettoyeur le chat y s'nettoie  
05 tout seul [m a i s c o m - ]  
06 K: [rien (.) moi::-]  
07 (0.24)  
08 J: Outre ça: j'veux dire toi t'achètes un animal de compagnie  
09 pour en prendre soin. Toi c'est bi[e::n t'as ton-]  
10 K: [J'en prends s]oi::n.  
11 (0.66)  
12 J: Oui::: mais c'est ça

La troisième catégorie, « après une discontinuité du tour », compte les chevauchements qui se produisent après un tour discontinu. Plus précisément, le tour de parole qui est chevauché n'est pas fluide. Ce manque de fluidité peut se caractériser par différents phénomènes comme la répétition de mots ou de syllabes, l'allongement d'une syllabe, un énoncé saccadé (entrecoupé de plusieurs courtes pauses) ou encore un ralentissement évident dans le débit de la parole. Les chevauchements apparaissent après discontinuité du tour précédent dans des proportions semblables pour chacun des contextes d'interaction, soit 10,32 % en présence, 8,22 % en vidéoconférence et 8,44 % en VR.

La quatrième catégorie, « au milieu du tour », correspond à des chevauchements qui débutent en l'absence d'un TRP apparent ou d'un silence, alors que le locuteur détient clairement le tour de parole (et qu'il est peu raisonnable de croire qu'il arrive à un TRP). Les chevauchements qui débutent au milieu d'un tour sont généralement des interruptions (le locuteur abandonne son tour avant d'arriver à une fin syntaxique et prosodique) ou encore des tentatives infructueuses de prise de parole par l'interactante (l'interactante abandonne son intervention après un mot ou une syllabe). En présence, 23,81 % des chevauchements débutent au milieu d'un tour alors qu'il y en a 16,67 % en vidéoconférence et 30,54 % en VR.

#### 5.3.2.2 Le type de chevauchement

Ce thème regroupe les principales caractéristiques qui sont présentes lors des chevauchements. Ces catégories sont : (1) les départs simultanés ; (2) les départs en canon ; (3) l'ajout du locuteur ; et finalement (4) les chevauchements qui impliquent un commentaire en lien direct avec l'environnement virtuel de la VR. Ces catégories ne sont pas exclusives. Par exemple, il est probable qu'un même chevauchement présente un *départ simultané* qui implique un *ajout du locuteur*. Nous avons retenu ces caractéristiques en raison de leur occurrence dans chacun des contextes individuellement. En d'autres mots, pour chacun des contextes d'interaction séparément, nous avons retenu les caractéristiques qui sont présentes dans plus de 10 % des chevauchements (parmi les chevauchements qui ne sont pas des BC).

La première caractéristique est le départ simultané. Comme leur nom l'indique, ces chevauchements impliquent la prise de parole des deux partenaires de conversation au même

moment. Les chevauchements marqués par un départ simultané se produisent généralement après un silence, alors que les interlocuteurs ont perdu le rythme de la conversation. Les départs simultanés surviennent dans 13,49 % des chevauchements en présence, 31,51 % de ceux en vidéoconférence et 15,79 % de ceux en réalité virtuelle. Il arrive aussi que les départs ne soient pas parfaitement simultanés : ce sont alors des départs *en canon*. Ces départs surviennent eux aussi principalement après un silence, mais la prise de parole est un peu décalée entre les partenaires : une personne débute un peu avant l'autre. Nos observations montrent des écarts qui varient entre 122 et 545 ms entre la prise de parole de chaque partenaire pour ce type de départ. Les départs en canon sont particulièrement nombreux dans les conversations en VR. Ils sont présents dans 18,95% des chevauchements en VR comparativement à 4,76 % de ceux en présence et 9,59 % de ceux en vidéoconférence.

#### **Extrait 4**

##### *Exemple d'un chevauchement en départ simultané*

01 K: Toi t'irais où?  
02 (1.99)  
03 K: [S'rait pas] à Moscou?  
04 J: [Moi : : :-]  
05 (0.59)

#### **Extrait 5**

##### *Exemple d'un chevauchement avec un départ décalé*

01 A: Fait que tu dois le promener avant d'aller à l'école (.) après  
02 qu'tu retournes d'l'école pi:s (.) pis ouains.  
03 (1.67)  
04 B: Ouais j'avoue [qu'ça prends quand mêm- (.) t'sais faut avoir le temps-]  
05 A: [Fait que est-ce t'as- (.) tu as le temps le matin]  
06 (0.69)  
07 B: Faut avoir le temps de s'occuper mais en même temps (.) c'est  
08 d'l'activité physique que com- qui s'rait quand même nice

L'Extrait 4 montre un exemple de départ simultané : les deux partenaires tentent de prendre la parole exactement au même moment après un long silence. Constatant que J ne répond pas à la



question de K posée à la ligne 01, K émet une hypothèse pour l'aider à répondre (ligne 03) au moment précis où J tente finalement d'émettre sa réponse (ligne 04). En comparaison, l'Extrait 5 montre aussi un chevauchement après un long silence, mais dans ce cas, le locuteur A amorce son énoncé un peu après B (ligne 05). Ces chevauchements sont des exemples d'un manque de coordination entre les partenaires de conversation menant à la violation du principe « sans silence ni chevauchement ». Le décalage du départ du chevauchement en VR semble impliquer la présence d'une plus grande latence dans ce contexte.

La troisième caractéristique est l'ajout par le locuteur. On entend ici qu'un locuteur ajoute un énoncé (complet ou non) à son tour de parole qui engendre un chevauchement avec le début du tour d'un autre locuteur. Ce type de chevauchement implique que la personne qui prend la parole le fait à un moment où il est raisonnable de croire que le locuteur initial a complété son tour de parole. Donc, ces chevauchements se produisent principalement après un TRP. Un exemple de ce type de chevauchement se trouve dans l'Extrait 6 où Z tente de prendre la parole (ligne 03) à un moment où Y aurait pu avoir terminé son tour de parole, mais Y se remet à parler (ligne 02). Ainsi, la reprise de la parole de Y à la suite d'un TRP engendre un court chevauchement. La proportion de chevauchements qui présente cette caractéristique est plutôt similaire d'un contexte à l'autre, soit 30,95 % en présence, 31,51 % en vidéoconférence et 20,00 % en VR.

### **Extrait 6**

#### *Exemple d'un chevauchement engendré par un ajout du locuteur*

01 Y: J'pense un chien ça fait pareil là- vraiment pareil (.)  
02 [Moi j']=  
03 Z: [Mai-]  
04 Y: =pense ça pas vraiment de différence entre les deux je sais  
05 pas pourquoi  
06 (1.00)  
07 Y: J' pense c' [est j]uste une histoire de- de:  
08 Z: [Moi-]  
09 (0.86)  
10 Y: Comme t'as dit

Finalement, la quatrième caractéristique fait référence aux chevauchements qui sont engendrés par un commentaire directement lié à l’environnement virtuel. Cette caractéristique est donc exclusive aux conversations en VR et elle correspond à 11,58 % des chevauchements. Les chevauchements de cette catégorie sont souvent des énoncés spontanés qui interrompent le flux de la conversation puisqu’ils arrivent habituellement au milieu d’un tour et qu’ils n’ont aucun lien avec ce qui est dit par le locuteur. Les commentaires font référence aux avatars, soit à cause de la manière dont ils bougent ou bien à cause d’un mouvement étrange (souvent un bras levé brusquement dans les airs) causé par un problème de reproduction entre la VR et la réalité. Dans l’exemple de l’Extrait 7, le locuteur H commence à parler (ligne 06) en chevauchement avec le locuteur G (ligne 05). Le locuteur H émet un commentaire directement relié à l’avatar du locuteur G parce que ses bras sont étrangement tordus derrière son dos. Ce problème est dû à la difficulté de la VR à reproduire correctement la position des bras lorsqu’ils sont croisés.

### **Extrait 7**

*Exemple d’un chevauchement directement associé au contexte de la réalité virtuelle*

```

01      G : Mais ouias c’est tellement [beau]=
02      H :                                     [okay]
03      G : =pis effectivement heu- plage montagne heu::
04          (0.54)
05      G : Vague heu:: (.) [ r i v i è r e   h e u : : : : ]
06      H :                                     [Qu’est-ce que tu fais avec tes bras]

```

#### **5.3.2.3 Le dénouement du chevauchement**

Finalement, le troisième thème utile pour décrire les chevauchements rassemble les événements qui se produisent après le chevauchement. Il y a trois catégories principales, soit (1) l’abandon du tour, (2) le silence et (3) la répétition. Ces événements ne sont pas exclusifs : un même chevauchement peut mener à la fois à l’abandon d’un énoncé, à un silence et à une répétition.

Premièrement, l’abandon du tour correspond aux chevauchements qui sont suivis par une interruption du discours de l’un des interlocuteurs (ou des deux) avant l’achèvement d’un énoncé syntaxique. Cet abandon permet de mettre fin le plus rapidement possible au chevauchement.

C'est le dénouement le plus fréquent dans le contexte en présence (61,11 %) et en vidéoconférence (84,93 %), alors qu'il est bon deuxième dans le contexte en VR (58,95 %).

Deuxièmement, les chevauchements sont souvent suivis d'un silence. Les silences se produisent immédiatement après le tour dans lequel le chevauchement s'est produit, soit lorsque les deux interlocuteurs abandonnent le tour simultanément, qu'ils achèvent tous deux leur énoncé ou encore une fois que le locuteur qui a conservé le tour termine son énoncé. Le silence est utilisé pour mettre fin au chevauchement en s'assurant que le canal de communication est libre avant de reprendre le cours de la conversation. Les chevauchements en VR sont davantage suivis de silences (60,00 %) que les chevauchements dans les contextes en présence (24,60 %) et en vidéoconférence (34,25 %). Encore une fois, le contexte en VR semble engendrer davantage de silences entre les partenaires de conversation que les autres contextes.

Troisièmement, les chevauchements sont parfois suivis de la répétition d'un élément qui est dit durant le chevauchement, au début de l'énoncé suivant. Cet élément peut être un mot, une syllabe ou encore un bout de phrase complet qui semble avoir été perdu dans le chevauchement et dont la répétition permet de compenser pour l'interférence causée par le chevauchement. La répétition est présente de façon similaire dans les différents contextes d'interaction, soit avec 11,90 % des chevauchements en présence, 10,96 % de ceux en vidéoconférence et 17,89 % de ceux en VR.

Finalement, d'autres phénomènes, dans des proportions moins importantes, sont observés à la suite des chevauchements : l'attribution du tour pour résoudre le chevauchement (par exemple, « Vas'y vas'y j't'écoute »), la demande de clarification (par exemple, « Hein ? ») ou la reconnaissance verbale qu'il y a un problème (par exemple, « J'ai pas compris »).

### **5.3.3 Conclusion sur le second objectif**

Pour répondre au second objectif de recherche, les événements problématiques semblent différents en VR et en vidéoconférence à plusieurs égards. D'abord, les silences en VR surviennent dans plus de contextes différents (ils ont des causes plus diversifiées) et il y a moins de gestes de communication qui sont produits durant ces silences pour maintenir l'interaction sociale entre les partenaires. Pour ce qui est des chevauchements, comparativement aux chevauchements en

vidéoconférence qui surviennent souvent en départ simultané, ceux en VR se produisent plus souvent avec un décalage. Cela dit, certaines similarités sont aussi observées. Entre autres, les chevauchements en VR et en vidéoconférence se produisent majoritairement après un silence et ils ont principalement le même dénouement. En effet, dans tous les contextes, les chevauchements sont suivis de l'abandon ou de l'interruption d'une ou des deux personnes, d'un silence et/ou d'une répétition au début du prochain énoncé.

## Chapitre 6 – Discussion

Dans la présente section, les résultats les plus saillants de la recherche sont interprétés afin de mieux comprendre les différences et les similarités entre les différents contextes d'interaction et ainsi répondre à la question de recherche de départ, à savoir : « comment les contextes d'interaction influencent-ils la coordination entre les partenaires de conversation? » Revoyons d'abord sommairement les résultats qui visent à répondre aux objectifs de cette recherche.

D'une part, les silences en VR sont différents que dans les autres contextes : ils sont plus longs. Effectivement, le contexte en VR contient davantage de longs silences que les contextes en présence et en vidéoconférence. Ces résultats confirment nos premières hypothèses (H1 et H2). Les résultats montrent aussi que les longs silences sont distribués différemment en VR que dans les autres contextes : en VR, il y a plus de longs intervalles que de longues pauses. Du côté des chevauchements, ces derniers semblent plus nombreux dans le contexte en VR qu'en vidéoconférence. Les comparaisons statistiques ne montrent pas d'évidence en faveur de l'hypothèse nulle ni en faveur de l'hypothèse alternative, donc plus de données sont nécessaires avant d'émettre des conclusions quant à notre troisième hypothèse (H3).

D'autre part, au point de vue qualitatif, les observations montrent des similarités et des différences entre les différents contextes d'interaction. Du côté des longs silences, ils sont surtout différents au niveau des causes apparentes et au niveau de l'interaction pendant le silence. Du côté des chevauchements, ces derniers ont plus de similarités : les chevauchements en VR et en vidéoconférence partagent des caractéristiques semblables quant à leur contexte d'apparition et à leur dénouement. Cependant, ils sont plus différents sur le plan du type de chevauchement.

Les prochaines sections explorent quatre explications possibles de ces résultats : la latence, la familiarité, les gestes et les expressions faciales, ainsi que la coprésence. Les silences et les chevauchements sont des phénomènes étroitement liés. C'est pour cette raison que les résultats sont regroupés en fonction de leurs interprétations possibles et non en fonction des indices de la coordination (silences et chevauchements), comme c'était le cas dans la section des résultats.

## 6.1 La latence

La latence est le temps qui est nécessaire au matériel technologique pour traiter et transmettre l'information d'une personne à une autre durant les conversations à distance. Il est attendu qu'une longue latence perturbe la fluidité des échanges qui ne sont pas en présence, tel que cela a été discuté au sujet des conversations en vidéoconférence (voir dans Seuren et al., 2021). Il semble raisonnable de croire que la VR implique un traitement de l'information plus complexe que celui de la vidéoconférence. La vidéoconférence ne fait que capter et reproduire des images et des sons à partir d'une caméra et d'un microphone. En comparaison, la VR doit produire des images animées (celles des avatars et de l'environnement virtuel) qui correspondent à l'information kinesthésique reçue (mouvement de la tête, du corps et des mains) en plus de transmettre l'information audio. D'ailleurs, la technologie requise pour utiliser la VR est beaucoup plus puissante que celle requise pour utiliser la vidéoconférence. On peut assister à des vidéoconférences sur pratiquement n'importe quel appareil électronique capable de se connecter à un réseau Internet : ordinateurs, téléphones intelligents et tablettes. Au contraire, la VR nécessite des ordinateurs de haute gamme, en raison de la complexité de la technologie. Comme le traitement et la transmission de l'information en VR est plus complexe qu'en vidéoconférence, la latence devrait être plus grande.

Une plus grande latence implique que le message est non seulement reçu plus lentement, mais aussi que sa réponse est transmise plus lentement. Donc, cette latence plus longue en VR pourrait expliquer la présence de silences plus longs en VR qu'en vidéoconférence. D'ailleurs, la latence permettrait aussi d'expliquer pourquoi, au point de vue qualitatif, les longs silences en VR surviennent régulièrement après un TRP. En VR, il y a plus de silences dont la cause apparente semble être simplement l'absence de prise de parole. Lorsque le locuteur complète son énoncé et qu'il n'y a pas de raison évidente de croire que le silence est dû à une réflexion, le long silence est possiblement la conséquence de la latence. Ainsi, un silence peut être perçu dans l'environnement virtuel même si en réalité, le locuteur B prend la parole avec un délai raisonnable après avoir perçu la fin de l'énoncé du locuteur A.

Par ailleurs, la latence pourrait aussi expliquer pourquoi les chevauchements en VR sont davantage émis avec des départs en canon qu'en départs simultanés. Il est possible que la personne qui commence à parler un peu après l'autre le fasse parce qu'elle n'a pas encore « entendu » l'autre. Dans ce cas, il semble raisonnable de supposer que ce genre de situation est évité en vidéoconférence puisque l'information est transmise plus rapidement (avec moins de latence). Autrement dit, la personne qui crée le chevauchement en VR éviterait le chevauchement en vidéoconférence parce qu'elle entendrait l'autre plus rapidement.

Un autre résultat que peut expliquer la latence est le nombre de chevauchements se produisant au milieu du tour de parole du locuteur, qui est plus élevé en VR qu'en présence et en vidéoconférence. Si une personne amorce son tour de parole en anticipant l'arrivée d'un TRP, mais que la personne qui détient le tour de parole le conserve après ce TRP, un chevauchement se produit. Ces chevauchements sont ceux qui se trouvent dans la catégorie d'apparition « près d'un TRP ». Il est possible que ces chevauchements soient perçus comme surgissant au milieu du tour du locuteur (plutôt que près d'un TRP) à cause de la latence. Ainsi, la personne qui émet le chevauchement le ferait (selon sa perception) près d'un TRP, mais à cause de la latence, il serait perçu plus tard, au milieu du tour. Ce type de problème se produit moins en présence et en vidéoconférence, ce qui est sans doute dû à l'absence de latence (ou à une latence moins grande) dans ces contextes.

## **6.2 La familiarité**

Les informations générales sur les individus ayant participé aux conversations (présentées à la section 5.1), indiquent qu'ils ont davantage utilisé la vidéoconférence que la VR par le passé. Donc, il est possible que les différences soulevées entre la vidéoconférence et la VR relèvent d'une meilleure adaptation des partenaires dans le premier contexte que dans le deuxième. Si les individus sont familiers avec le contexte en vidéoconférence, il est fort probable qu'ils aient déjà développé des habiletés à ajuster leurs comportements pour assurer une communication fluide malgré les contraintes du contexte.

La familiarité pourrait expliquer pourquoi les longs silences qui surviennent après un chevauchement se trouvent presque uniquement en VR (il y en a un seul en vidéoconférence et

aucun en présence). Tel que présenté dans les résultats qualitatifs des chevauchements, les silences surviennent fréquemment à la suite des chevauchements pour mettre fin au bris de communication et reprendre le rythme de la conversation. Donc les silences sont une manière habituelle de mettre fin aux chevauchements, mais le fait que ces silences soient longs uniquement en VR suggère que les chevauchements en VR sont plus difficiles à surmonter. Le fait qu'il y ait plus de longs silences après des chevauchements en VR est d'autant plus intéressant sachant que les conversations en présence contiennent plus de chevauchements que celles en VR (revoir le Tableau 4 à la section précédente).

D'un côté, la difficulté supplémentaire en VR pourrait s'expliquer par le fait que les partenaires sont moins bien habitués aux conditions de ce contexte d'interaction : ils n'arrivent pas à repérer les indices leur permettant de résoudre rapidement le chevauchement. En comparaison, les conversations en vidéoconférence n'utilisent presque pas les longs silences après les chevauchements. Il est possible que cela soit dû à leur plus grande familiarité avec ce contexte : les partenaires sont plus aptes à gérer leur bris de communication rapidement puisqu'ils sont plus à l'aise avec le contexte en vidéoconférence. D'un autre côté, les longs silences après les chevauchements pourraient plutôt être un signe d'adaptation : les partenaires s'ajustent en prenant plus de temps (des plus longs silences) après un chevauchement pour s'assurer que la voie audio est libre. Dans ce cas, le long silence serait utilisé comme un moyen efficace d'éviter de perpétuer davantage le chevauchement et permettrait ainsi de le résoudre plus rapidement. Le long silence serait alors un comportement adopté pour compenser les limites de la communication en VR.

Par ailleurs, il est attendu que les chevauchements dans les conversations soient nombreux (Levinson & Torreira, 2015). Pourtant, les résultats montrent qu'il y a moins de chevauchements en vidéoconférence qu'en présence et en VR. Cela peut s'expliquer par la difficulté apparente de résoudre les chevauchements en vidéoconférence comparativement à ceux en présence. Entre autres en raison du phénomène de latence déjà discuté, les partenaires ne perçoivent pas nécessairement le début du chevauchement en même temps. Donc, ils auront besoin de plus de temps pour le résoudre (Seuren et al., 2021). Si les chevauchements représentent un bris de communication plus important et plus difficile à surmonter en vidéoconférence, il est fort



probable que les partenaires cherchent à les prévenir le plus possible. Dans ce cas, il y aurait un effort supplémentaire en vidéoconférence pour ne pas simplement minimiser les chevauchements (comme en présence), mais plutôt pour les éviter complètement. Mais qu'en est-il de la VR ? Pourquoi les chevauchements en VR sont-ils plus nombreux qu'en vidéoconférence? Considérant que les chevauchements en VR semblent au moins aussi complexes à résoudre que ceux en vidéoconférence, on peut croire qu'il faudrait aussi chercher à les éviter en VR. Il est possible que les personnes produisent moins de chevauchements en vidéoconférence qu'en VR parce qu'ils ont plus l'habitude de discuter en vidéoconférence. Étant plus familiers avec la vidéoconférence que la VR, les partenaires arriveraient mieux à ajuster leurs comportements pour éviter les chevauchements dans ce contexte. C'est pourquoi il y aurait un plus grand nombre de chevauchements en VR.

### **6.3 Les gestes et les expressions faciales**

La VR limite la transmission des expressions faciales et peut transmettre différemment les gestes comparativement aux contextes en vidéoconférence et en présence. Ces différences au niveau visuel peuvent être liées aux longs silences de deux manières : l'absence de certains signaux visuels (les expressions faciales et la direction du regard) complique la gestion des tours de parole ce qui mène à de plus longs silences et/ou la transmission erronée d'un geste interrompt (ou distrait) le rythme de la conversation.

Par exemple, la fin d'un tour de parole est souvent accompagnée d'un contact visuel : le locuteur dirige son regard vers l'interactante pour lui signifier qu'elle peut prendre la parole (Kendon, 1967; Zima et al., 2019). En VR, la direction du regard doit être transmise par l'avatar. Une limite importante s'impose puisque les avatars peuvent ne pas avoir d'yeux, ou peuvent seulement avoir des yeux sans pupilles. De plus, même si l'avatar choisi avait des yeux et des pupilles qui bougent, ce mouvement des pupilles ne serait pas contrôlé par la personne qui l'incarne puisque la VR n'est pas équipée d'un système de suivi oculaire (*eye-tracker*). Par conséquent, il faut recourir aux mouvements grossiers de la tête pour interpréter la direction du regard d'une personne en VR, ce qui peut nuire à la finesse de la coordination entre les partenaires. En l'absence de regard précis, l'interactante risque d'avoir plus de difficulté à réaliser que le locuteur

lui cède la parole. Si l'interactante attend que le locuteur poursuive son tour alors que lui attend qu'elle prenne la parole, un long silence peut s'installer. Les contacts visuels en vidéoconférence ne peuvent pas être aussi francs que ceux en présence pour des raisons logistiques (la caméra qui capte le regard d'une personne est placée au-dessus de l'écran, donc au-dessus des yeux de l'interlocuteur et non en plein centre). Malgré tout, les partenaires peuvent avoir recours au mouvement des pupilles et mieux interpréter la direction du regard : il est possible de distinguer entre le fait de regarder droit devant (soit l'écran et donc l'interlocuteur) et le fait de regarder au-dessus ou en-dessous de l'écran.

Par ailleurs, durant les longs silences, des gestes de communication non-verbale comme un hochement de tête ou une expression faciale permettent d'assurer le maintien de l'interaction malgré l'absence de parole (les partenaires restent engagés activement dans la conversation). Si l'interaction est maintenue durant le long silence par un échange non-verbal entre les partenaires, il semble que ce long silence ne soit pas réellement l'effet d'un manque de coordination. Donc, ce sont davantage les longs silences durant lesquels l'interaction n'est pas maintenue qui semblent témoigner d'un manque de coordination. Encore une fois, il est possible que ces silences (ceux qui sont vides d'interaction) surviennent davantage en réalité virtuelle en raison de la plus grande latence du contexte en VR. Si le locuteur B reprend la parole après la fin du tour du locuteur A dans un délai qu'il considère raisonnable, alors il ne perçoit pas un long silence. Ainsi, il n'y aurait pas de gestes de communication durant le long silence puisque B ne perçoit pas de bris dans le rythme de la conversation qui pourrait l'amener à émettre un comportement pour en assurer le maintien.

Lorsqu'il y a un silence, le rythme de la conversation peut être perdu, ce qui peut mener à un chevauchement entre les partenaires lors de la reprise de la parole. Les mouvements (comme la posture et l'ouverture de la bouche pour se préparer à parler) peuvent faciliter la reprise de la parole fluide à la suite des silences. Le fait que la transmission de l'information visuelle en VR soit limitée pourrait expliquer qu'on observe davantage de chevauchements à la suite des silences en VR qu'en vidéoconférence. Les partenaires peuvent s'appuyer sur davantage de repères en vidéoconférence qu'en VR (tel que le montre la plus grande présence de gestes

communicationnels durant les longs silences en vidéoconférence qu'en VR) pour arriver à un consensus quant à l'attribution du tour de parole à la suite du silence.

L'absence de certains indices visuels en VR peut expliquer pourquoi les chevauchements près d'un TRP sont moins nombreux dans ce contexte qu'en vidéoconférence. Les chevauchements qui se produisent près d'un TRP supposent que l'interactante ait tenté d'anticiper la fin du tour du locuteur pour prendre la parole au TRP. Même si cette anticipation mène à un chevauchement, l'événement peut tout de même être signe que le contexte offre assez d'informations aux partenaires pour assurer une certaine justesse dans l'anticipation<sup>13</sup> (et par le fait même dans la coordination). Donc, s'il y a moins de chevauchements près d'un TRP dans le contexte en VR, ce pourrait être dû au manque d'indices pragmatiques (expressions faciales, gestes, etc.) qui limite la capacité des partenaires à anticiper l'arrivée d'un TRP. En l'absence d'indices adéquats pour anticiper les TRP, les partenaires ne prennent la parole que lorsqu'ils sont certains que le tour est complété, évitant par le fait même de produire un chevauchement. Dans le même ordre d'idées, les partenaires oseraient davantage prendre la parole en anticipant la fin d'un tour, au risque de provoquer un chevauchement, lorsqu'ils sont en présence (contexte dans lequel les chevauchements près des TRP sont plus nombreux) que dans les contextes à distance.

## 6.4 La coprésence

La coprésence, ou présence sociale (de l'anglais *social presence*), fait référence au degré avec lequel les partenaires de conversation se sentent en proximité et engagés dans l'interaction (Cui et al., 2013). L'engagement dans la conversation est lié à la fréquence des chevauchements : lorsque les partenaires sont plus engagés dans la conversation, celle-ci est plus dynamique avec moins de silences et possiblement plus de chevauchements (Lestary et al., 2018).

Il est possible de supposer que le sentiment de coprésence en VR est plus fort qu'en vidéoconférence et ressemble davantage à celui en présence. En effet, le caractère immersif de la VR permet aux partenaires d'être complètement submergés (du moins au niveau visuel) dans

---

<sup>13</sup> Rappelons d'autant plus que les chevauchements qui se produisent après des TRP sont généralement engendrés par l'ajout du locuteur à son tour précédent qui peut difficilement être anticipé (même en contexte naturel) par l'interactante. Dans ce cas, la bonne coordination du tour de parole de l'interactante n'est pas nécessairement mise en cause.

l'environnement virtuel. Cette immersion donne l'impression d'être physiquement au même endroit que l'autre comme en face-à-face. En comparaison, le sentiment de coprésence est probablement plus faible en vidéoconférence puisque les partenaires sont conscients, sans l'ombre d'un doute, qu'ils ne se trouvent pas au même endroit ; l'environnement qu'ils *partagent* est limité à leur écran d'ordinateur. Pour résumer, s'il y a presque autant de chevauchements en VR qu'en présence (soit plus qu'en vidéoconférence), c'est peut-être parce que les interlocuteurs sont plus engagés dans la conversation grâce au sentiment de coprésence.

Cependant, on pourrait aussi argumenter que le sentiment de coprésence est plus faible en VR qu'en vidéoconférence et en présence. Cela serait dû au fait que l'interaction passe par le biais d'avatars et que l'information visuelle est réduite aux gestes grossiers de la tête et du corps. Comme les personnes ne voient pas réellement leur interlocuteur, mais plutôt l'image d'un personnage animé plus ou moins humanoïde, il est probable que les partenaires ressentent moins de proximité dans la conversation, l'avatar agissant comme une barrière. Ainsi, bien qu'elles partagent un même environnement virtuel, les personnes pourraient être moins portées à s'engager dans les conversations en VR par comparaison à celles en vidéoconférence, où les partenaires voient l'image réelle de leur interlocuteur. Cela pourrait expliquer le fait qu'il y a une plus faible proportion de silences de réflexion en VR que dans les autres contextes.

Les silences de réflexion, qui sont les plus récurrents dans tous les contextes d'interaction, sont attendus dans le cadre de cette recherche étant donné les procédures expérimentales utilisées (les partenaires doivent répondre à des questions qui portent sur des thèmes précis pendant des conversations d'environ cinq minutes) – c'est d'ailleurs pourquoi nous associons la forte présence de ce type de silences au type de tâche de discussion, et non aux contextes d'interaction à l'étude. La proportion plus faible des silences de réflexion en VR pourrait s'expliquer par le faible sentiment de coprésence des partenaires : puisqu'ils se trouvent face à un avatar et non à un humain, ils ressentiraient moins le besoin de s'engager dans la conversation, réfléchiraient moins à ce qu'ils disent et porteraient leur attention sur autre chose que la conversation. De fait, en plus des réflexions, les longs silences en VR sont apparemment causés par un facteur qui ne se trouve ni dans le contexte en présence ni dans le contexte en vidéoconférence : les distractions propres à la VR. Ainsi, il est possible que le contexte en VR ne soit pas associé à un plus fort sentiment de

coprésence. Au contraire, la VR qui implique l'utilisation d'avatars pourrait plutôt nuire au sentiment de coprésence et à l'engagement dans la conversation, ce qui aurait un effet négatif sur la fluidité de la communication.

## **6.5 Conclusion de la discussion**

De ces principaux résultats quantitatifs et qualitatifs, on peut conclure que le contexte en VR influence la coordination (les silences et les chevauchements) dans la conversation de manière différente que la vidéoconférence. En effet, les silences dans les conversations en VR sont plus longs et plus nombreux, ce qui confirme nos premières hypothèses. De plus, les silences et les chevauchements en VR surviennent dans des contextes différents et ils se déroulent différemment. Toutes ces différences entre la VR et la vidéoconférence pourraient s'expliquer, entre autres, par la latence plus grande en VR, la plus grande familiarité des individus avec le contexte en vidéoconférence, les limites de la VR dans la transmission de l'information visuelle fine et le sentiment de coprésence qui se manifeste différemment dans les différents contextes.

## Chapitre 7 – Conclusion

Notre étude est née du questionnement suivant : en quoi la coordination en réalité virtuelle est-elle différente de celle en vidéoconférence ? Au terme de notre étude, nous pouvons conclure que dans le contexte en VR, par rapport à la présence et à la vidéoconférence, les conversations semblent moins fluides : la coordination entre les partenaires est plus difficile. En VR, la durée moyenne des silences est plus longue que dans les autres contextes d'interaction et il y a un plus grand nombre de longs silences. De plus, nos données tendent à montrer qu'il y a davantage de chevauchements en VR qu'en vidéoconférence. Par ailleurs, nos analyses qualitatives permettent d'approfondir notre compréhension de l'effet des contextes d'interaction sur la coordination. Entre autres, les longs silences ne semblent pas partager les mêmes causes et les types de chevauchements semblent différents d'un contexte à l'autre. Nous avons développé ces résultats en discutant de quatre explications possibles : la latence, la familiarité, les gestes et les expressions faciales ainsi que la coprésence. Si notre interprétation est juste, ces quatre éléments sont des facteurs qui influencent différemment la coordination en VR et en vidéoconférence. Il existe probablement d'autres facteurs, mais les résultats de notre étude offrent un bon point de départ pour les recherches futures.

### 7.1 Les limites de la recherche et réflexions pour les recherches futures

Dans ce mémoire, nous nous intéressons à un nouveau contexte d'interaction : la réalité virtuelle. Conséquemment, nous avons fait face à des limites qui doivent être considérées pour l'interprétation de nos résultats et pour les recherches futures.

Au niveau de la quantité de données recueillies, des recherches ultérieures pourraient valider nos observations avec un échantillon plus large. Cela dit, au niveau des analyses qualitatives, les données recueillies nous permettent d'atteindre la saturation des données : les 240 silences et les 293 chevauchements observés semblent nous permettre d'avoir une idée complète de l'ensemble du phénomène d'intérêt, c'est-à-dire les situations où la conversation n'est pas fluide.

Pour notre étude, nous avons documenté le niveau de familiarité des dyades avec les différents contextes d'interaction dans un questionnaire d'informations générales, mais nous n'avons pas étudié en profondeur ce facteur. Comme la VR gagne en popularité et devient plus accessible à la population générale, il sera de plus en plus possible pour les recherches futures de recruter des personnes ayant un niveau élevé de familiarité avec la VR. Par conséquent, ces recherches pourront se concentrer davantage sur l'effet de la familiarité avec les contextes d'interaction sur la gestion des tours de parole dans la conversation. Par exemple, une étude pourrait comparer les contextes d'interaction avec des dyades ayant un niveau de familiarité équivalent pour les deux contextes (en vidéoconférence et en VR). Une autre étude pourrait comparer des dyades qui sont très familières avec les contextes à celles qui sont peu familières avec les contextes.

De plus, les futures recherches pourraient inclure une période de familiarisation avec le contexte en VR plus longue dans leur protocole. Dans notre étude, les partenaires avaient environ 10 à 15 minutes durant l'installation du matériel pour se familiariser avec l'environnement virtuel. Une période de familiarisation plus longue permettrait de diminuer l'effet de nouveauté du contexte, permettant possiblement de réduire les distractions qui y sont associées. Par ailleurs, en contrôlant pour le niveau de familiarité avec le contexte, les recherches futures pourraient distinguer les comportements de communication qui semblent être causés par un manque d'adaptation au contexte des comportements qui semblent plutôt être utilisés comme moyens pour pallier les limites des contextes d'interaction à distance.

Finalement, la technologie en matière de réalité virtuelle progresse rapidement. Par conséquent, il existe déjà du matériel plus performant que celui utilisé pour la collecte des données de la présente étude. Par exemple, nous savons que des systèmes permettant de transmettre la direction du regard ainsi que les mouvements de la bouche sont nouvellement disponibles (voir le casque *Meta Quest Pro* inauguré au mois d'octobre 2022). Il va sans dire que nos résultats abondent en faveur du perfectionnement de ces technologies qui permettront de modérer les limites que nous avons observées en VR. Cela dit, ces nouveaux systèmes sont très coûteux et ils ne sont donc pas aussi accessibles à la population générale que le matériel utilisé dans notre recherche. À plus long terme, il sera intéressant d'étudier la façon dont ces progrès

technologiques améliorent réellement (ou non) la communication en VR comparativement à ce que nous observons pour l'instant.

## **7.2 Les retombées pour le domaine de la pratique**

Avec le développement des différentes technologies de communication, les professionnels de la santé – médecins, psychologues, orthophonistes, pour ne nommer que ceux-là – peuvent maintenant offrir des consultations à distance à leurs patients. La crise sanitaire causé par la COVID-19 a mené au confinement général de la population durant une période prolongée de temps. Ce faisant, les consultations à distance ont été davantage favorisées. La meilleure compréhension des éléments en jeu dans les différents contextes d'interaction à distance peut permettre aux professionnels de faire des choix plus judicieux quant à la manière de communiquer avec leur patient.

À titre d'exemple, si un professionnel souhaite utiliser la réalité virtuelle avec un patient (pour l'attrait de la nouveauté et de l'aspect immersif de l'expérience en VR par exemple), il doit être conscient des difficultés de communication qui peuvent être engendrées par ce choix : ses interactions avec son patient risquent d'être moins fluides, de comporter davantage de longs silences et des distractions. Le professionnel doit aussi être conscient des limites associées à la communication non-verbale en VR et choisir judicieusement le matériel qu'il utilisera. Tel que nos résultats le montrent, les équipements de VR qui sont les plus accessibles pour l'instant ne permettent pas de transmettre le déplacement des yeux des avatars indiquant précisément la direction du regard des interlocuteurs. Ceci est une limite importante de la VR, puisque plusieurs informations pragmatiques sont transmises par le regard dans les conversations. Le professionnel qui souhaite utiliser la VR peut se procurer du matériel plus performant qui (nous l'espérons) permettra de réduire les effets que nous avons observés, ou encore utiliser du matériel plus accessible en s'assurant d'adapter ses comportements communicationnels en conséquence (par exemple, en utilisant davantage des mouvements grossiers de la tête et des mains pour montrer la direction de son regard et pour donner la parole à son interlocuteur).

En conclusion, nos résultats soulèvent l'existence de différences entre les contextes de communication à distance qui suggèrent que l'anticipation permettant la coordination fluide des



échanges de tour de parole est plus limitée dans le contexte en VR que dans celui en vidéoconférence. Est-ce en raison du medium lui-même (la VR n'est tout simplement pas propice à la fluidité), à cause de la technologie (la VR n'est pas suffisamment développée) ou à cause des utilisateurs de ce medium (ils ne savent pas l'utiliser de telle sorte qu'ils aient des conversations fluides) ? Les futures études dans le domaine pourront se concentrer sur ces questions précises et éventuellement développer suffisamment de connaissances pour élaborer des procédures pour guider les personnes désirant intégrer la VR à leurs pratiques en recherche et en clinique.

## Références bibliographiques

- Adams, C. (2002). Practitioner review: the assessment of language pragmatics. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 43(8), 973-987. <https://doi.org/10.1111/1469-7610.00226>
- Airenti, G. (2017). Pragmatic Development. Dans L. Cummings (dir.), *Research in Clinical Pragmatics* (pp. 3-28). Springer International Publishing. [https://doi.org/10.1007/978-3-319-47489-2\\_1](https://doi.org/10.1007/978-3-319-47489-2_1)
- Annahita, M. E., & Mary, J. D. S. (2015). Social capital and common mental disorder: a systematic review. *Journal of Epidemiology and Community Health*, 69(10), 1021. <https://doi.org/10.1136/jech-2015-205868>
- Aubin-Auger, I., Mercier, A., Baumann, L., Lehr-Drylewicz, A.-M., Imbert, P., & Letrilliart, L. (2008). Introduction à la recherche qualitative. *Exercer*, 84(19), 142-145.
- Auzanneau, M., & Boutet, J. (2021). Interaction. *Langage et société, Hors série(HS1)*, 179-183. <https://doi.org/10.3917/l.s.hs01.0180>
- Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27(2), 195-225. [https://doi.org/10.1016/S0364-0213\(02\)00118-0](https://doi.org/10.1016/S0364-0213(02)00118-0)
- Barthel, M., & Levinson, S. C. (2020). Next speakers plan word forms in overlap with the incoming turn: evidence from gaze-contingent switch task performance. *Language, Cognition and Neuroscience*, 35(9), 1183-1202. <https://doi.org/10.1080/23273798.2020.1716030>
- Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition*, 203, 104347. <https://doi.org/10.1016/j.cognition.2020.104347>
- Bögels, S., & Torreira, F. (2021). Turn-end Estimation in Conversational Turn-taking: The Roles of Context and Prosody. *Discourse Processes*, 58(10), 903-924. <https://doi.org/10.1080/0163853X.2021.1986664>
- Bohannon, L. S., Herbert, A. M., Pelz, J. B., & Rantanen, E. M. (2013). Eye contact and video-mediated communication: A review. *Displays*, 34(2), 177-185. <https://doi.org/10.1016/j.displa.2012.10.009>

- Boland, J. E., Fonseca, P., Mermelstein, I., & Williamson, M. (2022). Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General*, 151, 1272-1282. <https://doi.org/10.1037/xge0001150>
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, 44(2), 123-147. <https://doi.org/10.1177/00238309010440020101>
- Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, É., Forget, H., Loranger, C., & Roucaut, F. X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *The British Journal of Psychiatry*, 210(4), 276-283. <https://doi.org/10.1192/bjp.bp.116.184234>
- Bouchard, S., St-Jacques, J., Robillard, G., & Renaud, P. (2007). Efficacité d'un traitement d'exposition en réalité virtuelle pour le traitement de l'arachnophobie chez l'enfant une étude pilote. *Journal de Thérapie Comportementale et Cognitive*, 17(3), 101-108. [https://doi.org/10.1016/S1155-1704\(07\)73238-X](https://doi.org/10.1016/S1155-1704(07)73238-X)
- Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. *Language and Speech*, 37(1), 1-20. <https://doi.org/10.1177/002383099403700101>
- Chowdhury, S., Stepanov, E., Danieli, M., & Riccardi, G. (2017). Functions of Silences towards Information Flow in Spoken Conversation. Dans N. Ruiz & S. Bangalore (Eds.), *Proceedings of the First Workshop on Speech-Centric Natural Language Processing* (pp. 1-9). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-4601>
- Clancy, P. M., Thompson, S. A., Suzuki, R., & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese and Mandarin. *Journal of Pragmatics*, 26(3), 355-387. [https://doi.org/10.1016/0378-2166\(95\)00036-4](https://doi.org/10.1016/0378-2166(95)00036-4)
- Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation. *Discourse Processes*, 55(2), 230-240. <https://doi.org/10.1080/0163853X.2017.1330031>
- Cosnier, J. (2016). Les Gestes du dialogue. Dans *La Communication* (pp. 112-121). Éditions Sciences Humaines. <https://doi.org/10.3917/sh.dorti.2016.02.0112>

- Croes, E. A. J., Antheunis, M. L., Schouten, A. P., & Krahmer, E. J. (2019). Social attraction in video-mediated communication: The role of nonverbal affiliative behavior. *Journal of Social and Personal Relationships, 36*(4), 1210-1232. <https://doi.org/10.1177/0265407518757382>
- Croteau, C., McMahon-Morin, P., Le Dorze, G., Power, E., Fortier-Blanc, J., & Davis, G. A. (2018). Exploration of a quantitative method for measuring behaviors in conversation. *Aphasiology, 32*(3), 247-263. <https://doi.org/10.1080/02687038.2017.1350629>
- Cui, G., Lockee, B., & Meng, C. (2013). Building modern online social presence: A review of social presence theory and its instructional design implications for future trends. *Education and Information Technologies, 18*(4), 661-685. <https://doi.org/10.1007/s10639-012-9192-1>
- de la Rosa, S., & Breidt, M. (2018). Virtual reality: A new track in psychological research. *British Journal of Psychology, 109*(3), 427-430. <https://doi.org/10.1111/bjop.12302>
- de Ruiter, J. P., & Albert, S. (2017). An Appeal for a Methodological Fusion of Conversation Analysis and Experimental Psychology. *Research on Language and Social Interaction, 50*(1), 90-107. <https://doi.org/10.1080/08351813.2017.1262050>
- Degutyte, Z., & Astell, A. (2021). The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings. *Frontiers in Psychology, 12*, 903. <https://www.frontiersin.org/article/10.3389/fpsyg.2021.616471>
- Dittmann, A. T., & Llewellyn, L. G. (1968). Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology, 9*(1), 79-84. <https://doi.org/10.1037/h0025722>
- Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied, 3*(2), 105-125. <https://doi.org/10.1037/1076-898X.3.2.105>
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology, 23*(2), 283-292. <https://doi.org/10.1037/h0033031>
- Duncan, S., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory*. Lawrence Erlbaum Associates.

- Eiswirth, M. E. (2020). Increasing interactional accountability in the quantitative analysis of sociolinguistic variation. *Journal of Pragmatics*, 170, 172-188. <https://doi.org/10.1016/j.pragma.2020.08.018>
- Ephratt, M. (2008). The functions of silence. *Journal of Pragmatics*, 40(11), 1909-1938. <https://doi.org/10.1016/j.pragma.2008.03.009>
- Finch, E., Cameron, A., Fleming, J., Lethlean, J., Hudson, K., & McPhail, S. (2017). Does communication partner training improve the conversation skills of speech-language pathology students when interacting with people with aphasia? *Journal of Communication Disorders*, 68, 1-9. <https://doi.org/10.1016/j.jcomdis.2017.05.004>
- Fox Tree, J. E. (2002). Interpreting Pauses and Ums at Turn Exchanges. *Discourse Processes*, 34(1), 37-55. [https://doi.org/10.1207/S15326950DP3401\\_2](https://doi.org/10.1207/S15326950DP3401_2)
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Prentice-Hall.
- Hamad, A., & Jia, B. (2022). How Virtual Reality Technology Has Changed Our Lives: An Overview of the Current and Potential Applications and Limitations. *International Journal of Environmental Research and Public Health*, 19(18), 11278. <https://www.mdpi.com/1660-4601/19/18/11278>
- Hayashi, M., Raymond, G., & Sidnell, J. (2013). Conversational repair and human understanding: an introduction. Dans G. Raymond, J. Sidnell, & M. Hayashi (Eds.), *Conversational Repair and Human Understanding* (pp. 1-40). Cambridge University Press. <https://doi.org/10.1017/CBO9780511757464.001>
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555-568. <https://doi.org/10.1016/j.wocn.2010.08.002>
- Hepburn, A., & Potter, J. (2021). *Essentials of conversation analysis*. American Psychological Association.
- Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, 25(5), 1900-1908. <https://doi.org/10.3758/s13423-017-1363-z>

- Hömke, P., Holler, J., & Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLOS ONE*, *13*(12), e0208030. <https://doi.org/10.1371/journal.pone.0208030>
- Ignatowicz, A., Atherton, H., Bernstein, C. J., Bryce, C., Court, R., Sturt, J., & Griffiths, F. (2019). Internet videoconferencing for patient–clinician consultations in long-term conditions: A review of reviews and applications in line with guidelines and recommendations. *DIGITAL HEALTH*, *5*, 2055207619845831. <https://doi.org/10.1177/2055207619845831>
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101-144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Jakobson, R. (1960). Concluding statement: linguistics and poetics. Dans T. A. Sebeok (dir.), *Style in language* (pp. 350-377). Technology Press of Massachusetts Institute of Technology.
- JASP Team. (2023). *JASP (Version 0.17.2)[Computer software]*. Dans <https://jasp-stats.org/>
- Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Human Studies*, *9*(2), 153-183. <https://doi.org/10.1007/BF00148125>
- Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. Dans *Conversation: An interdisciplinary perspective*. (pp. 166-196). Multilingual Matters.
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773-795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, *26*, 22-63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- Kendrick, K. H. (2017). Using Conversation Analysis in the Lab. *Research on Language and Social Interaction*, *50*(1), 1-11. <https://doi.org/10.1080/08351813.2017.1267911>
- Kendrick, K. H., & Holler, J. (2017). Gaze Direction Signals Response Preference in Conversation. *Research on Language and Social Interaction*, *50*(1), 12-32. <https://doi.org/10.1080/08351813.2017.1262120>
- Kerbrat-Orecchioni, C. (1998). *Les Intéractions verbales* (3e ed.). Armand Colin.

- Khademi, S. (2022). A Review of Silence in Conversation: Discoursal Perspective. *International Journal of Language and Translation Research*, 2(1), 51-67. [https://doi.org/10.12906/9783899664812\\_004](https://doi.org/10.12906/9783899664812_004)
- Knudsen, B., Creemers, A., & Meyer, A. S. (2020). Forgotten Little Words: How Backchannels and Particles May Facilitate Speech Planning in Conversation? [Original Research]. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.593671>
- Kogure, M. (2007). Nodding and smiling in silence during the loop sequence of backchannels in Japanese conversation. *Journal of Pragmatics*, 39(7), 1275-1289. <https://doi.org/10.1016/j.pragma.2007.02.011>
- Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, 73, 1-40. <https://doi.org/10.1016/j.cogpsych.2014.04.001>
- Koudenburg, N., Postmes, T., & Gordijn, E. H. (2011). Disrupting the flow: How brief silences in group conversations affect social needs. *Journal of Experimental Social Psychology*, 47(2), 512-515. <https://doi.org/10.1016/j.jesp.2010.12.006>
- Koudenburg, N., Postmes, T., & Gordijn, E. H. (2013). Conversational Flow Promotes Solidarity. *PLOS ONE*, 8(11), e78363. <https://doi.org/10.1371/journal.pone.0078363>
- Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology*, 43(4), 718-731. <https://doi.org/10.1037/0022-3514.43.4.718>
- Lestary, A., Krismanti, N., & Hermaniar, Y. (2018). Interruptions and Silences in Conversations: A Turn-Taking Analysis. *Parole: Journal of Linguistics and Education*, 7(2), 53-64. <https://doi.org/10.14710/parole.v7i2.64>
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 731. <https://www.frontiersin.org/article/10.3389/fpsyg.2015.00731>
- Lu, Y., Ge, Y., Chen, W., Xing, W., Wei, L., Zhang, C., & Yang, Y. (2022). The effectiveness of virtual reality for rehabilitation of Parkinson disease: an overview of systematic reviews with meta-analyses. *Systematic Reviews*, 11(1), 50. <https://doi.org/10.1186/s13643-022-01924-5>

- Magyari, L., & de Ruiter, J. (2012). Prediction of Turn-Ends Based on Anticipation of Upcoming Words. *Frontiers in Psychology*, 3, 376. <https://www.frontiersin.org/article/10.3389/fpsyg.2012.00376>
- Malbos, E., Boyer, L., & Lançon, C. (2013). L'utilisation de la réalité virtuelle dans le traitement des troubles mentaux. *La Presse Médicale*, 42(11), 1442-1452. <https://doi.org/10.1016/j.lpm.2013.01.065>
- McFarland, D. H. (2001). Respiratory Markers of Conversational Interaction. *Journal of Speech, Language, and Hearing Research*, 44(1), 128-143. <https://pubs.asha.org/doi/abs/10.1044/1092-4388%282001/012%29>
- Monninger, M., Aggensteiner, P.-M., Pollok, T. M., Kaiser, A., Reinhard, I., Hermann, A., Reichert, M., Ebner-Priemer, U. W., Meyer-Lindenberg, A., Brandeis, D., Banaschewski, T., & Holz, N. E. (2023). The importance of high quality real-life social interactions during the COVID-19 pandemic. *Scientific Reports*, 13(1), 3675. <https://doi.org/10.1038/s41598-023-30803-9>
- Ono, E., Nozawa, T., Ogata, T., Motohashi, M., Higo, N., Kobayashi, T., Ishikawa, K., Ara, K., Yano, K., & Miyake, Y. (2011, 20-22 Décembre). Relationship between social interaction and mental health. 2011 IEEE/SICE International Symposium on System Integration (SII), Kyoto, Japan.
- Orange, J. B., Van Gennep, K. M., Miller, L., & Johnson, A. M. (1998). Resolution of communication breakdown in dementia of the Alzheimer's type: A longitudinal study. *Journal of Applied Communication Research*, 26(1), 120-138. <https://doi.org/10.1080/00909889809365495>
- Pan, X., & Hamilton, A. F. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3), 395-417. <https://doi.org/10.1111/bjop.12290>
- Park, M. J., Kim, D. J., Lee, U., Na, E. J., & Jeon, H. J. (2019). A Literature Overview of Virtual Reality (VR) in Treatment of Psychiatric Disorders: Recent Advances and Limitations [Mini Review]. *Frontiers in Psychiatry*, 10. <https://doi.org/10.3389/fpsyg.2019.00505>



- Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review*, 26(3), 894-900. <https://doi.org/10.3758/s13423-019-01571-3>
- Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: mechanisms and information sources. *Frontiers in Psychology*, 6(89). <https://doi.org/10.3389/fpsyg.2015.00089>
- Robert, A. D., & Bouillaguet, A. (2007). Chapitre II. Méthodologie générale de l'analyse de contenu et application à un exemple. Dans *L'analyse de contenu* (pp. 24-46). Presses Universitaires de France. <https://www.cairn.info/l-analyse-de-contenu--9782130563839-page-24.htm>
- Ross, S. (2018). Listener response as a facet of interactional competence. *Language Testing*, 35(3), 357-375. <https://doi.org/10.1177/0265532218758125>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696-735. <https://doi.org/10.2307/412243>
- Salehi, A., Ehrlich, C., Kendall, E., & Sav, A. (2019). Bonding and bridging social capital in the recovery of severe mental illness: a synthesis of qualitative research. *Journal of Mental Health*, 28(3), 331-339. <https://doi.org/10.1080/09638237.2018.1466033>
- Schegloff, E. A. (2001). Accounts of Conduct in Interaction: Interruption, Overlap, and Turn-Taking. Dans J. H. Turner (dir.), *Handbook of Sociological Theory* (pp. 287-321). Springer US. [https://doi.org/10.1007/0-387-36274-6\\_15](https://doi.org/10.1007/0-387-36274-6_15)
- 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. <https://doi.org/10.2307/413107>
- Schutz, S., Walthall, H., Snowball, J., Vagner, R., Fernandez, N., Bartram, E., & Merriman, C. (2022). Patient and clinician experiences of remote consultation during the SARS-CoV-2 pandemic: A service evaluation. *DIGITAL HEALTH*, 8, 20552076221115022. <https://doi.org/10.1177/20552076221115022>
- Seedhouse, P. (2022). Transitioning from conversation analysis to mixed methods. *Language Teaching*, 1-12. <https://doi.org/10.1017/S0261444822000027>

- Sellen, A. J. (1995). Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, 10, 401-444. [https://doi.org/10.1207/s15327051hci1004\\_2](https://doi.org/10.1207/s15327051hci1004_2)
- Seuren, L. M., Wherton, J., Greenhalgh, T., & Shaw, S. E. (2021). Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, 172, 63-78. <https://doi.org/10.1016/j.pragma.2020.11.005>
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587. <https://doi.org/10.1073/pnas.0903616106>
- ten Bosch, L., Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, 47(1), 80-86. <https://doi.org/10.1016/j.specom.2005.05.009>
- Tolins, J., & Fox Tree, J. E. (2014). Addressee backchannels steer narrative development. *Journal of Pragmatics*, 70, 152-164. <https://doi.org/10.1016/j.pragma.2014.06.006>
- van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., Dablander, F., Gronau, Q. F., Kucharský, Š., Gupta, A. R. K. N., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.-J. (2020). A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP. *L'Année psychologique*, 120(1), 73-96. <https://doi.org/10.3917/anpsy1.201.0073>
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209-232. <https://doi.org/10.1016/j.specom.2013.09.008>
- Westfall, P. H., Johnson, W. O., & Utts, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika*, 84(2), 419-427. <https://www.jstor.org/stable/2337467>
- Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, 12(6), 957-968. <https://doi.org/10.3758/BF03206432>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation, Gênes, Italie.

- Xu, J. (2014). Displaying status of reciprocity through reactive tokens in Mandarin task-oriented interaction. *Journal of Pragmatics*, 74, 33-51. <https://doi.org/10.1016/j.pragma.2014.08.008>
- Yngve, V. H. (1970). On Getting a Word in Edgewise. Dans *Papers from the Sixth Regional Meeting, Chicago Linguistic Society* (pp. 567-578). Chicago Linguistic Society.
- Zima, E., Weiß, C., & Brône, G. (2019). Gaze and overlap resolution in triadic interactions. *Journal of Pragmatics*, 140, 49-69. <https://doi.org/10.1016/j.pragma.2018.11.019>

## Annexe A

**Tableau 7**

*Listes des tâches de conversation par paires*

Paires de tâches	Dilemmes <sup>a</sup>	Thèmes de conversation <sup>b</sup>
1	Marie et Paul sont à l'hôpital. Ils trouvent 100\$ sur le plancher. Que feriez-vous à la place de Marie et Paul?	Quel est le meilleur contexte de travail pour vous et pourquoi? (En présence ou en télétravail)
2	Marie et Paul sont couchés. Ils entendent des bruits de pas dans la maison. Ils sont inquiets. Que feriez-vous à la place de Marie et Paul?	Quel est le meilleur animal de compagnie selon vous et pourquoi?
3	Marie et Paul voient des jeunes devant la maison des voisins. Ils essaient d'ouvrir la porte de leur maison. Que feriez-vous à la place de Marie et Paul?	Quel est votre destination de voyage idéale et pourquoi?
4	Marie et Paul jouent aux cartes. Ils croient avoir vu un de leurs amis tricher. Que feriez-vous à la place de Marie et Paul?	Dans quel type de quartier aimeriez-vous vivre plus tard et pourquoi?
5	Marie et Paul font une excursion en forêt. Ils se perdent. Que feriez-vous à la place de Marie et Paul?	Selon vous, quels gestes quotidiens permettent de mieux protéger l'environnement et pourquoi?

<sup>a</sup> Mises en situation de Croteau et al. (2018). <sup>b</sup> Questions inspirées de McFarland (2001).