



**HAL**  
open science

## Multimodalité(s) dans les interactions humaines

Mary Amoyal, Marco Cappellini, Christelle Combe, Berthille Pallaud,  
Béatrice Priego-Valverde, Marion Tellier

► **To cite this version:**

Mary Amoyal, Marco Cappellini, Christelle Combe, Berthille Pallaud, Béatrice Priego-Valverde, et al.. Multimodalité(s) dans les interactions humaines. Travaux Interdisciplinaires sur la Parole et le Langage, 2022, Numéro spécial : Panorama des recherches au Laboratoire Parole et Langage, 38. hal-03948653

**HAL Id: hal-03948653**

**<https://hal.science/hal-03948653v1>**

Submitted on 20 Jan 2023

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

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

# TIPA. Travaux interdisciplinaires sur la parole et le langage

38 | 2022

Numéro spécial : Panorama des recherches au Laboratoire  
Parole et Langage

---

## Multimodalité(s) dans les interactions humaines

*Multimodality in human interactions*

Mary Amoyal, Marco Cappellini, Christelle Combe, Berthille Pallaud,  
Béatrice Priego-Valverde et Marion Tellier

---



Édition électronique

URL : <https://journals.openedition.org/tipa/5167>

ISSN : 2264-7082

Éditeur

Laboratoire Parole et Langage

---

Ce document a été généré automatiquement le 20 janvier 2023.

---

# Multimodalité(s) dans les interactions humaines

*Multimodality in human interactions*

Mary Amoyal, Marco Cappellini, Christelle Combe, Berthille Pallaud, Béatrice Priego-Valverde et Marion Tellier

---

## Introduction

- 1 Dans notre contribution, nous allons aborder la notion de multimodalité dans les interactions humaines. Dans un premier temps, nous mettrons au jour les différentes définitions du concept de multimodalité actuellement utilisées dans les recherches en sciences du langage. Dans un deuxième temps, nous détaillerons certaines recherches actuelles sur la multimodalité menées au LPL en montrant la variété des cadres théoriques et méthodologiques mobilisés ainsi que des situations d'interaction considérées, allant du face à face présentiel aux interactions en ligne, de la conversation entre locuteurs d'une même langue à la communication exolingue. Enfin, dans une dernière partie, nous montrerons que cette variété n'empêche pas la complémentarité, ni des ressources langagières mobilisées par les participants, ni des cadres théoriques et méthodologiques qui en rendent compte. Nous concluons cet article en proposant quelques pistes de recherche future notamment à partir du phénomène de défaillance interactionnelle.

## 1. Le concept de multimodalité

- 2 La communication interpersonnelle (Goffman, 1974) est « multimodale » dans son acception la plus large, c'est-à-dire en prenant en compte aussi bien les travaux de recherche menés en sciences du langage sur l'oral (Colletta, 2004), sur la gestuelle (Cosnier *et al.* 1982, McNeill, 1992), en analyse conversationnelle (Mondada, 2005) que ceux de la sémiotique sociale (Kress & Van Leeuwen, 2001) ou encore de la communication médiée par ordinateur (Develotte *et al.*, 2011) et de l'analyse du

discours numérique (Paveau, 2017). Dans cette première partie, nous définissons nos différents ancrages théoriques principaux, l'un n'excluant pas l'autre.

## 1.1. Eclairages linguistiques

- 3 Lorsque l'on analyse la parole en interaction en tant que phénomène multimodal, on considère en général trois grandes modalités : le verbal, le vocal (dont la prosodie [Voir l'article "La prosodie au Laboratoire Parole et Langage : histoire, recherches actuelles et perspectives"<sup>1</sup> dans ce numéro]) et les éléments posturo-mimo-gestuels.

« La parole est multimodale. [...] [Elle] est un comportement qui associe des sons articulés à des mouvements corporels, et lorsque vous percevez une conduite langagière, vous percevez celle-ci dans les différentes modalités sensorielles dont vous êtes équipés. Autrement dit, non seulement vous reconnaissez les mots prononcés, mais vous les entendez avec leur musique (ce qu'on appelle la prosodie), et vous percevez en même temps les gestes, les mimiques, les regards...l'ensemble des mouvements corporels produits par le locuteur. » Colletta (2004 : 15).

- 4 Cette définition de la parole multimodale proposée par Colletta met en évidence les différentes modalités en présence dans la perception de la parole à l'oral, en face à face. Elle peut être complétée par la définition de Ferré (2011 : 74) :

« L'analyse multimodale consiste à mettre en relation des informations linguistiques produites dans différentes modalités, chacune d'elle contribuant à l'élaboration et à la perception du message communiqué. Ainsi, l'on peut distinguer la modalité verbale qui comporte plusieurs niveaux — phonèmes, choix du lexique, organisation syntaxique, organisation discursive — de la modalité orale — prosodie, qualité de voix — et enfin de la modalité visuelle — gestualité et expressions faciales. »

- 5 La définition de Ferré, quant à elle, se place du point de vue de l'analyse. Une linguistique multimodale prend donc en compte les différentes modalités et les analyse en les mettant en relation. Étudier la prosodie ou analyser le geste (ou d'autres aspects kinésiques tels que les postures, les regards et les mimiques faciales) sans le verbal isolerait les modalités et ne permettrait pas de montrer leur articulation.
- 6 Les interactions en face à face impliquent l'utilisation de plusieurs modalités langagières parmi lesquelles la parole et les expressions faciales. Lorsque l'on considère la communication comme « un tout intégré » (Winkin, 1981) et non comme une opposition entre communication verbale et non verbale, une analyse multimodale est tout à fait pertinente. De nombreuses études se sont intéressées au rôle des « gestes faciaux » (Bavelas *et al.*, 2014) en lien avec la parole produite telles que des études sur les regards (Kendrick & Holler, 2017), les mouvements de sourcils (Goujon, 2019), les clignements des yeux (Homke *et al.*, 2018), les hochements de tête (Stivers, 2008). Parmi ces différentes expressions faciales, le sourire est un geste très peu étudié dans son rapport avec ce qui est dit par les interlocuteurs. Le sourire a principalement été étudié en perception (Aubergé & Cathiard, 2003 ; Fagel, 2010) ou en relation avec ses fonctions émotionnelles (Ekman, 1975 ; Tartter, 1980). Pourtant, le sourire est un comportement social destiné à la communication : il est déterminé et mieux prédit par le contexte social plutôt que par les émotions intérieures (Fridlund, 1994). Dans cet article nous présenterons la méthodologie utilisée pour décrire les liens entre sourire et discours produits dans des interactions en face à face.

## 1.2. Eclairage de la sémiotique sociale

- 7 Une conception de la multimodalité par l'entrée des modalités sensorielles peut être complexifiée si l'on pense qu'une même modalité peut être le canal de plusieurs modes sémiotiques. Ainsi, à travers la modalité visuelle, parfois appelée canal de la vue, il est possible de percevoir plusieurs modes sémiotiques : expressions faciales, gestes et proxémique entre autres. Ainsi, la multimodalité peut être définie comme une caractérisation de l'interaction non seulement en termes de modalités à l'œuvre, mais aussi de *modes sémiotiques*.
- 8 Le développement des technologies de l'audio-visuel et du numérique a fait émerger dans les années 1990, une autre vision du concept de multimodalité, ancrée dans le champ de la sémiotique sociale. Les travaux de Kress et van Leeuwen (2001) ont permis de définir de nouveaux contours de cette notion en intégrant les différentes modalités de la communication permise par les différents artefacts utilisés. Ainsi la multimodalité est définie comme le recours massif et conjoint à divers modes d'expression (verbal, visuel, audio, tactile, etc.) dans la communication. Dans le champ de la sémiotique sociale, d'autres concepts sont généralement associés à la multimodalité tel que le canal (qui est l'entrée sensorielle d'après les cinq sens humains - vue, ouïe, tact, goût, odorat) ; la modalité (qui est l'entrée d'après l'artefact utilisé - par exemple, pour la visioconférence, on compte principalement les modalités visuelle et auditive) et le mode (qui correspond à un régime sémiotique mobilisé dans des ressources sémiotiques et en fonction de leurs affordances, autrement dit les possibilités et contraintes d'un environnement pour un acteur engagé dans une action). Dans ce cadre, selon Bezemer et Kress (2016) une ressource sémiotique sera définie comme une entité mixte liant les moyens (matériels ou immatériels) et les modes sémiotiques. Des exemples de moyens matériels sont le papier comme support à l'écriture, au dessin, etc. Les moyens non matériels sont principalement les actions, les relations entre interlocuteurs, les genres discursifs.

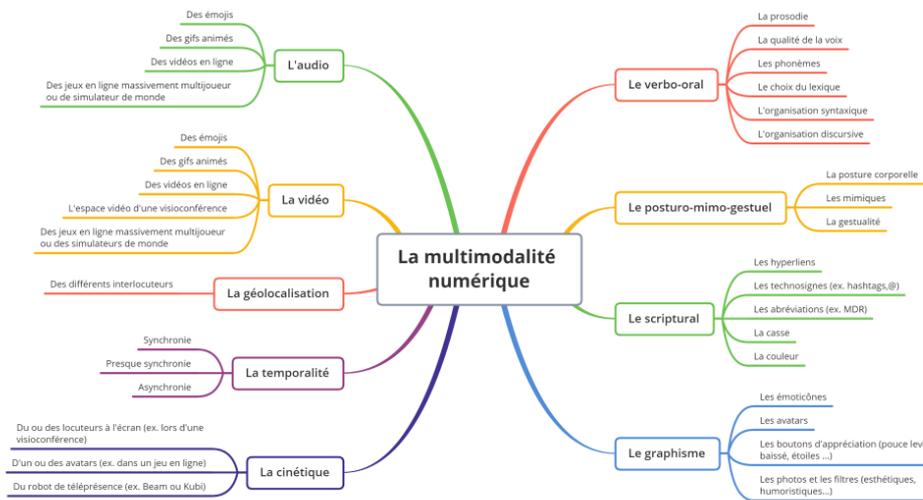
## 1.3. Eclairages émergents dans la « multimodalité numérique »

- 9 Le concept de « multimodalité numérique » (Wachs & Weber, 2021) s'ancre dans plusieurs champs de recherches, proches les uns des autres mais présentant chacun des particularités. Le premier, fondateur et précurseur est le champ de la Communication Médinée par Ordinateur (CMO) qui s'est intéressé, dès le début d'Internet (et même du minitel en France) aux échanges en ligne, principalement sous leur forme textuelle (Herring, 1996 ; Anis, 1998). Viennent ensuite deux champs émergents de la linguistique dans la lignée de la CMO : le champ des Interactions Multimodales Par Ecran<sup>2</sup> (IMPEC) qui s'est développé notamment autour des interactions par visioconférence (Develotte, Kern & Lamy, 2011, Develotte & Paveau, 2017 ; Guichon & Tellier, 2017) et celui de l'Analyse du Discours Numérique (Paveau, 2017) qui consiste en « la description et l'analyse du fonctionnement des productions langagières natives d'internet, et plus particulièrement du web 2.0, dans leurs environnements de production, en mobilisant à considération égale les ressources langagières et non langagières des énoncés élaborés » (*ibid.* : 27). Enfin, dans le domaine de la didactique des langues, on soulignera le champ de l'Apprentissage des Langues médiatisé par la Technologie (Guichon, 2012)

qui étudie l'intégration des technologies dans les situations de communication spécifiquement liées à l'enseignement-apprentissage des langues.

- 10 Présent donc, dès lors que deux modes se sont trouvé mêlés, comme le textuel et le graphique à travers les premiers émoticônes par exemple, le concept de « multimodalité » en contexte numérique est devenu de plus en plus central : ainsi, Herring (2015) parle désormais de « Communication interactive multimodale » (Interactive Multimodal Communication) comprenant la communication multimodale médiée par ordinateur (principalement sur les plateformes interactives multimodales) et la communication médiée par robot (Robot-Mediated Communication). Il est aussi devenu de plus en plus complexe. En effet, si originellement la CMO reposait sur des éléments textuels que l'on étudiait au sein des courriels, forums de discussion, groupes de diffusion, donjons multi-utilisateurs, blogues, microblogues et wikis, désormais le graphique, l'audio, la vidéo sont de plus en plus disponibles sur les plateformes du web 2.0 et les smartphones (Herring & Androutsopoulos, 2015). La CMO est considérée comme multimodale quand sa production ou sa réception implique un canal de communication autre ou en addition au textuel (Herring & Androutsopoulos, 2015). On peut citer les émoticônes, les dessins animés (emoji), les gifs animés, les avatars, les bannières, les tickers, les boutons de relationnalité, les mèmes internet, les pancartes numériques, les incrustations textuelles, les reprofilages par filtres, l'ensemble de ce que Paveau (2017) nomme « technographismes ». Les annotations vidéo collaboratives (CVA) sont des variantes dynamiques de la communication textuelle au sein de l'image ainsi que les commentaires faits dans un mode autre que textuel. Comme Herring et Androutsopoulos (2015) le soulignent la CMO exclusivement textuelle est en déclin, la multimodalité est de plus en plus forte (Develotte, Kern & Lamy, 2011), les frontières entre les codes s'estompent ou se reconfigurent dans les écosystèmes connectés (Paveau, 2017), on tend vers une « iconisation du texte » (Paveau, 2019) grandissante, le métissage des signes (scripturaux, sonores, iconiques et datas) devenant la norme.
- 11 Parallèlement, du point de vue de la formation des individus et des compétences des citoyens à l'ère numérique, Lacelle, Boutin et Lebrun (2018 : 8) définissent la littératie médiatique multimodale appliquée en contexte numérique comme : « la capacité d'une personne à mobiliser adéquatement, en contexte communicationnel synchrone ou asynchrone, les ressources et les compétences sémiotiques modales (ex : mode linguistique seul) et multimodales (ex : combinaison des modes linguistique, visuel et sonore) les plus appropriées à la situation et au support de communication (traditionnel et (ou) numérique), à l'occasion de la réception (décodage, compréhension, interprétation et évaluation) et (ou) de la production (élaboration, création, diffusion) de tout type de message ».
- 12 Ainsi, étudier la communication multimodale numérique consiste donc à mettre en relation des informations linguistiques, sémiotiques et technodiscursives produites dans différentes modalités chacune d'elle contribuant à l'élaboration et à la perception du message communiqué dans une communauté donnée. Le schéma ci-dessous synthétise l'ensemble des éléments que le chercheur peut prendre en compte dans son analyse de la multimodalité numérique dans la mesure où ceux-ci seront croisés.

Figure 1 : La multimodalité numérique



- 13 Sur le plan méthodologique, la transcription de la multimodalité numérique s'avère également un enjeu pour les chercheurs. Guichon (2013) propose un exemple de transcription multimodale à partir de la thèse de Drissi (2011) transcrivant l'ensemble des modalités présentes : l'audio, le tchat et la webcam. Par ailleurs, avec les possibilités multimodales qu'offre désormais l'édition numérique, la transcription de la communication multimodale numérique peut elle-même exploiter ce champ des possibles.

## 2. Les recherches sur la multimodalité au LPL

- 14 Les chercheur.e.s du LPL qui s'intéressent à la question de la multimodalité étudient chacun des contextes différents comme les interactions en face-à-face, les interactions en ligne et les interactions hybrides.

### 2.1. Les interactions en face-à-face

- 15 Les interactions en face-à face peuvent être de différentes nature (e.g. conversations, interaction médecin-patient, réunion de travail...). Dans cette section, nous présenterons certaines recherches qui portent sur la conversation en face-à-face. Nos travaux portent principalement sur la prosodie lors de disfluences et sur le sourire dans deux phases de la conversation : les transitions thématiques et les phases d'humour.

#### 2.1.1. Interactions conversationnelles

- 16 Les études actuelles menées au LPL s'inscrivent dans les cadres de l'Analyse Conversationnelle (Sacks, Schegloff, & Jefferson, 1974) et de la Linguistique Interactionnelle (Couper-Kuhlen & Selting, 1996). A ce titre, la conversation est perçue comme une « production collaborative » (Sacks, Schegloff, & Jefferson, 1974) pendant laquelle les interlocuteurs mobilisent plusieurs « ressources multimodales » (Birdwhistell, 1968).

### **Sourire et conversation**

- 17 Parmi les nombreuses ressources mobilisées, nous nous intéressons plus particulièrement aux sourires que les interlocuteurs déploient au cours d'une conversation. S'il est admis que les signes du visage sont particulièrement observés par l'interlocuteur (Barrier, 2013), il a également été montré que le sourire du locuteur contribue à maintenir l'attention de son interlocuteur (Theonas, Hobbs, & Rigas, 2008). Dès lors, nous appréhendons le sourire comme un « geste facial » (Bavelas *et al.*, 2014) participant à la collaboration des interlocuteurs au cours d'une conversation. Quelques fonctions interactives du sourire ont déjà été analysées comme le sourire-feedback, le sourire de politesse ou le sourire de requête (Koop *et al.*, 2008, Argyle, 1988). Au LPL, nous mettons l'accent sur le rôle du sourire dans deux moments-charnières de la conversation au cours desquels les interlocuteurs négocient un changement potentiel de trajectoire interactionnelle (Wooffitt, 2005) : la transition thématique (changement de thème conversationnel) et l'humour (changement de registre, ou de « frame », Priego-Valverde *et al.*, 2018).
- 18 Le sourire fait intervenir plusieurs muscles du visage tels que (entre autres) les zygomatiques et le muscle orbiculaire de l'œil. La sollicitation de ces différents muscles et les différences observées entre un sourire faible ou fort montrent qu'il est important de prendre en compte les différentes intensités des sourires (El haddad, 2019). Ces évidences nous ont amené à aborder le sourire en fonction des différentes intensités déployées par les interlocuteurs et non plus de manière binaire (présence/absence). En effet, la seule présence d'un sourire ne permet pas de rendre compte de la diversité de ses fonctions (i.e. un sourire faible peut se voir attribuer une fonction différente par rapport à un large sourire). C'est pourquoi, notre méthodologie d'annotation prend appui sur l'échelle d'intensités de sourire (Smiley Intensity Scale, SIS, (Gironzetti, Attardo, & Pickering, 2016) qui distingue trois intensités de sourires (sourire bouche fermée S1, sourire bouche ouverte S2, sourire grande bouche ouverte S3). Par ailleurs, cette échelle permet d'annoter le visage neutre (S0) et le rire (S4). Cette méthodologie de description graduelle du sourire permet de s'intéresser à l'évolution des sourires mais également à la réciprocité des sourires des interlocuteurs. Afin de réaliser une annotation des sourires sur de larges corpus, nous avons mis en place un protocole d'annotation semi-automatique des sourires [cf. l'article "Principes et outils pour l'annotation des corpus"<sup>3</sup> dans ce numéro].
- 19 Afin de réaliser une analyse multimodale de conversations en français, nous étudions des corpus audio-visuels de conversations semi-contrôlées. Dans les deux corpus étudiés « Cheese!<sup>4</sup> » (Priego-Valverde, Bigi, & Amoyal, 2020) et PACO<sup>5</sup> (Amoyal, Priego-Valverde, & Rauzy, 2020 (cf. fiches techniques), les interlocuteurs commencent par lire à tour de rôle des histoires drôles puis posent les textes pour démarrer une discussion libre. Le signal de parole est ensuite transcrit et les sourires annotés [cf. l'article "Principes et outils pour l'annotation des corpus"<sup>6</sup> dans ce numéro] afin de pouvoir ensuite analyser ces différentes modalités de concert. Enfin, les transitions thématiques et les occurrences d'humour sont également annotées [cf. l'article sur l'annotation dans ce numéro : "Principes et outils pour l'annotation des corpus"<sup>7</sup>].

### *Sourire et transition thématique*

- 20 À l'appui de la transcription, une analyse interactionnelle est menée afin de décrire l'enchaînement thématique des conversations. Les transitions thématiques sont ensuite étudiées dans leur rapport avec les sourires des interlocuteurs. Une analyse mixte de nos données est menée mêlant d'une part des analyses descriptives et statistiques de nos données afin de dégager les tendances de configurations de sourires et, d'autre part, des analyses séquentielles des transitions de nos deux corpus qui, combinés avec les tendances dégagées, permettront de mettre en avant les processus linguistiques mis en place par les participants dans les différentes phases des transitions analysées.
- 21 Cette analyse a montré plusieurs résultats intéressants.
1. La fin d'un thème est majoritairement associée à un sourire.
  2. Les locuteurs changent systématiquement leur intensité de sourire lorsqu'ils effectuent une transition thématique.
  3. Ce changement observé est une baisse du sourire des locuteurs qui présentent ainsi un visage neutre en rupture avec le sourire précédent la transition. Cette configuration (i.e. suppression du sourire) se retrouve dans 30 % des transitions vs 9 % dans le reste de la conversation.
  4. Cette suppression du sourire pour changer de thème est d'autant plus fréquente pour des interlocuteurs qui ne se connaissent pas (PACO) vs. des interlocuteurs qui se connaissent bien (Cheese!).
- 22 Nous allons illustrer ces résultats à l'appui d'un extrait du corpus PACO (Corpus dans lequel les interlocuteurs se rencontrent pour la première fois : cf. fiche technique<sup>8</sup>).
- 1 LB : (S2) je j'en fais le plus possible donc euh voilà (S4) @
- 2 LE : (S2) mh
- 3 LB : (S2) je suis là pour ça
- 4 → LE : (S2) ok (S0) mais t'es étudiant à \$Marseille\$ ou à \$Aix\$
- 5 LB : (S0) à \$Aix\$
- 6 LE : (S0) oui d'accord ok ouais
- 23 Cet extrait apparaît juste après que les interlocuteurs ont expliqué chacun la raison de leur présence mutuelle à l'expérience en cours. LB est le dernier à avoir exprimé la raison de sa présence donc à alimenter le thème précédent. A la L.1, il conclue cette séquence thématique en expliquant qu'il réalise le plus d'expérience que possible, puis il émet plusieurs marqueurs de clôture thématique « donc euh voilà » ponctué par un léger rire, ces éléments combinés étant des marqueurs de la fin de son explication. Puis il résume son intervention en L.3 en reformulant le thème abordé « je suis là pour ça ». Ces deux énoncés sont prononcés avec un S2 alors que LB ne souriait pas juste avant ces énoncés. Le FB en L.2 de LE permet à LB de poursuivre son énoncé en ayant la confirmation que son interlocutrice suit son discours et ainsi de conserver son rôle interactionnel de locuteur. En L.4, LE prend le tour de parole avec le feedback « ok » qui collabore en montrant qu'elle a compris ce qui a été dit avec un S2. LE propose un nouveau thème à l'appui d'une question fermée autour de leur lieu d'étude. Ce changement de thème introduit par la conjonction de coordination « mais » est de type « enchaînement » puisqu'il est lié au thème précédent dans la mesure où LE demande une précision sur son lieu d'étude pour mieux comprendre sa présence à l'expérience en cours. En effet, LE a expliqué la raison de sa présence à l'expérience par le fait que

son lieu d'étude est identique au lieu de la passation et de ce fait projette cette raison sur son interlocuteur. Par ailleurs, cette proposition est une question ce qui permet à LE de conserver sa position d'interlocutrice en invitant LB à développer. Cette proposition est accompagnée par une suppression du sourire puisque le début de l'énoncé est réalisé avec un S2 puis un visage neutre. En L.5, LB répond à la question posée avec un visage neutre et accepte ainsi le nouveau thème qui sera co-développé pendant 35 secondes.

### **Sourire et humour**

- 24 A ce jour, peu de travaux analysent l'activité mimo-gestuelle des interlocuteurs comme une *ressource parmi d'autres* pour « faire de l'humour ». Parmi ces travaux, la manière dont le locuteur humoriste indique à son partenaire qu'il fait ou qu'il va faire de l'humour – autrement dit, qu'il change de registre – est la piste la plus explorée. Ainsi, Tabacaru et Lemmens (2014) ont montré le rôle des mouvements de sourcils pour indiquer ce passage à un registre humoristique. Les mouvements de sourcils fonctionneraient donc comme des « gestural triggers ». L'étude a été certes conduite sur des séries télévisées, mais elle offre des pistes de recherche intéressantes et mériterait d'être appliquée sur des corpus conversationnels. Par ailleurs, les travaux de Gironzetti (2017) et Gironzetti *et al.* (2016) sur l'activité mimo-gestuelle (plus spécifiquement le sourire et le regard) des interlocuteurs lors de séquences humoristiques sont probablement les plus aboutis. Ainsi, concernant le sourire, Gironzetti (2017) apporte les deux résultats suivants : (1) locuteur et interlocuteur sourient avec une intensité plus grande juste avant, pendant et juste après les moments humoristiques. (2) Les locuteurs et interlocuteurs produisent plus fréquemment des sourires réciproques, et de même intensité lors des séquences humoristiques.
- 25 Dans la lignée des travaux menés par Gironzetti (2017) et Gironzetti *et al.* (2016), les recherches (encore en cours) du LPL sur l'implication du sourire dans un énoncé humoristique sont menées dans deux directions complémentaires. La première (Priego-Valverde *et al.*, 2018) relève d'une comparaison interculturelle de la manière dont les interlocuteurs sourient. Sur la base d'un corpus en anglais-américain et d'un autre en français (« Cheese! »), il s'agissait ici d'observer l'utilisation du sourire par les interlocuteurs selon qu'ils produisent ou non de l'humour. Il a alors été montré le même impact de l'humour sur la présence de sourires dans les deux corpus : les interlocuteurs déploient davantage de sourires réciproques et de même intensité lors de moments d'humour. En revanche, ils déploient davantage de visages neutres, ou de sourires d'intensités différentes lors de moments sérieux.
- 26 Une étude exploratoire plus récente (Priego-Valverde *et al.* 2020) conduite uniquement sur deux interactions du corpus « Cheese! » a confirmé le rôle du sourire comme ressource utilisée par le locuteur pour indiquer qu'il bascule dans l'humour. Par ailleurs, la présence vs. absence de sourire d'une part, et son niveau d'intensité d'autre part, ont été analysés dans deux conditions : selon que l'énoncé humoristique réussisse ou échoue. Les premières analyses permettent de dégager deux tendances : (1) l'intensité des sourires produits lors d'un énoncé humoristique qui réussit est plus élevée que lors d'un énoncé humoristique qui échoue, (2) même lorsque l'humour échoue, les interlocuteurs produisent davantage de sourires que lors des moments sérieux de la conversation. Il semblerait donc que le sourire n'empêche pas l'échec de l'humour.

### *Disfluences et conversation*

- 27 Des marqueurs de clôture thématique tels que « donc euh voilà » se trouvent impliqués dans les analyses de discours sur la transition thématique (changement de thème conversationnel) et les études sur l'humour (changement de registre, ou de « frame »). Or, ces marqueurs sont également, comme on va le voir, des marqueurs impliqués dans les nombreux cas de rupture de la fluence verbale lors des conversations, la transition thématique n'étant qu'un cas parmi d'autres.
- 28 Tout comme la seule présence d'un sourire ne permet pas de rendre compte de la diversité de ses fonctions, les disfluences présentes dans les énoncés oraux des locuteurs ne permettent pas d'augurer ni de leurs fonctions ni de leur contexte multimodal. Qu'ils soient produits lors de conversations, de monologues ou de conférences, ils comportent de nombreuses variations de la fluence verbale et, cela, à plusieurs niveaux (par exemple, le débit de prononciation des mots, des syntagmes ou des propositions (Pasdeloup, 1992 ; Duez, 2001b ; Shriberg, 1994). Si on envisage ces phénomènes de fluence verbale d'un point de vue multimodal, on se rend compte que ces variations se manifestent à la fois sur le plan acoustique, phonétique telles que la prosodie, les phénomènes de conglomérats ou de réductions stéréotypés (Cole *et al.* 2005 ; Shriberg, 1995 ; Bertrand *et al.*, 2008) mais également sur les plans morphologique et syntaxique (c'est-à-dire verbal) Sur ces plans morphologique et syntaxique, certaines de ces variations se traduisent par de véritables *auto-interruptions*<sup>9</sup> qui suspendent le déroulement syntagmatique dans l'émission verbale. Les conventions de transcription des corpus du CID<sup>10</sup> (Corpus of Interactional Data ; Bertrand *et al.*, 2008), ont prévu de conserver les traces d'élaboration des énoncés que sont, entre autres, les amorces ou fragments de mots et les ruptures de syntagme. Cette stratégie a permis d'envisager une description fine et exhaustive de ces phénomènes désignés sous le terme de *disfluence*. Les travaux entrepris dès 1999 au LPL (Pallaud, 2015) ont adopté la description de la structure de ces phénomènes proposée par Shriberg (1994) : le *Reparandum* désignant l'élément interrompu à réparer ou abandonner, suivi de l'*Interregnum* pour ce qui suit l'interruption sans intervention sur le *Reparandum*, et ensuite le *Reparans* qui correspond à la modification de l'élément interrompu. Les phénomènes de disfluences ont été décrits et annotés selon cette typologie et ont donné lieu à l'annotation exhaustive des corpus du CID [cf. l'article "Principes et outils pour l'annotation des corpus"<sup>11</sup> dans ce numéro] : dans l'exemple suivant, le segment *complèt-* (le *Reparandum*) est le résultat d'une interruption marquée à la fois sur le plan prosodique et phonétique ; il sera suivi immédiatement (donc sans espace *Interregnum*) d'une reprise (le *Reparans*) jusqu'au début de la phrase contenant le segment interrompu *je suis complètement partagé*.

Ex 1 ; AG-1027.260 **je complèt-***je suis complètement partagé*

- 29 Les espaces *Interregnum*, nommés ainsi parce qu'ils ne participent pas à la construction syntaxique interrompue, sont situés entre ce que Shriberg a nommé *Reparandum* (l'élément qui précède l'interruption et *Reparans* l'élément qui reprend la construction syntaxique interrompue). Comme on vient de le voir, les *Interregnum* peuvent être vides (//) ou contenir des éléments oraux (pauses remplies ou non, éléments discursifs, interjections, termes variés) qui sont insérés dans l'énoncé tels des incises non reliées syntaxiquement au texte oral. C'est le cas de l'exemple suivant :

Ex 2 ; AG-983.284 ça contribue à //à **si tu veux** à cette //à cette construction

30 L'analyse morphosyntaxique de ces disfluences a montré que le *Reparandum* pouvait être un fragment de mot (Ex 1) ou un mot ou un syntagme (Ex 2). De plus, les disfluences se subdivisent en trois catégories selon que le *Reparans* répète ou complète l'élément interrompu sans modification (Ex 2 ; cas les plus fréquents), ou qu'il le modifie (Ex 3 ; un quart des disfluences). Plus rarement encore, le locuteur laisse le mot ou la construction syntaxique inachevés (Ex 4) :

Ex 3 ; AG-1047.381 J'ai euh moi j'ai lev- // j'ai lu un truc sur ça

Ex 4 ; AG-1425.850 Tu te trompes un peu de mot et c'est // l'expression elle veut plus rien dire

31 Les disfluences sont donc parties intégrantes de l'élaboration de l'énoncé et sont si fréquentes qu'elles occupent en moyenne un tiers de la durée totale de l'énoncé (Pallaud & Bertrand, 2020). Plusieurs facteurs peuvent les favoriser :

- Le vieillissement : plus on avance en âge, plus on tend à produire des discours disfluents (Bolly *et al.*, 2016) ;
- Le contexte de prise de parole (publique par rapport à la parole spontanée) : les marqueurs de discours qui sont très liés aux disfluences, sont eux aussi très représentatifs d'une planification en élaboration. Leur présence peut varier fortement et montre une augmentation croissante avec la prise de parole spontanée et non préparée (par opposition à la prise de parole planifiée telle qu'elle peut l'être en public) (Beliao & Lacheret, 2013) ;
- Certaines pathologies semblent également plus sensibles à la situation de parole spontanée : dans les cas d'hyperactivité (Engelhardt, 2010) ou d'autisme (Lake *et al.*, 2011), les disfluences dans les énoncés se révèlent être plus fréquentes.

32 Les effets des disfluences ont été étudiés surtout dans le cadre d'apprentissage :

- Les disfluences signalent les hésitations du locuteur. Perçues par le récepteur, elles accroissent son attention et la mémorisation du mot contexte de cette disfluence (Collard *et al.*, 2008) ;
- La situation d'apprentissage de langue seconde (Nivja de Jong, 2018) : dans une situation d'apprentissage de langues, la durée des pauses semble être plus révélatrice du style de l'apprenant qu'une spécificité de la situation d'apprentissage. Par contre, le débit de parole et son inverse (la durée moyenne de prononciation des syllabes) sont caractéristiques de l'apprenant L2.

33 En somme, les études sur les disfluences effectuées au LPL se sont concentrées sur l'analyse du verbal et des paramètres acoustico-prosodiques. Il nous semblerait tout à fait pertinent d'intégrer les gestes co-verbaux dans cette analyse (cf. partie Perspectives).

### 2.1.2. Interactions exolingues

34 Les interactions exolingues, que Porquier définit comme une communication établie « par des moyens autres qu'une langue maternelle éventuellement commune aux participants » (Porquier, 1994 : 18), sont particulièrement intéressantes à étudier d'un point de vue multimodal. La communication exolingue que nous étudions relève de deux types : les interactions dans un milieu institutionnel (la classe de langue ou l'enseignement en ligne synchrone, voir ci-dessous) et les interactions en contexte semi-contrôlé (dans une tâche proposée à des participants de langues différentes). Dans chacune des situations, il y a un expert de la langue (enseignant expérimenté ou en formation) et un apprenant. Pour reprendre Py, « l'apprentissage d'une langue, ainsi

que la communication exolingue qu'elle permet se définissent tous deux par un effort vers la construction d'une sorte de rationalité et d'intelligibilité linguistique. » (2000 : 401). Bange (1992) a conceptualisé cette attention qui dans la communication exolingue est portée à la fois sur le contenu et sur l'intelligibilité du contenu par tous les interlocuteurs en termes de bi-focalisation. Bien entendu, des adaptations verbales et vocales sont mises en œuvre par l'enseignant ou le locuteur expert pour faciliter la compréhension et l'acquisition (entre autres, Ferguson, 1975 ; Long, 1983). En outre, bien que moins étudiée, l'utilisation du corps, et du geste en particulier, vient souvent servir de béquille voire même prendre le relais de la parole pour faciliter l'accès au sens (Tellier, 2008). En analysant comment les discours sont construits multimodalement en situation exolingue, nous avons mis au jour plusieurs phénomènes en interaction en face à face.

- 35 Une première analyse concerne la façon dont le geste et la parole s'articulent. On observe trois différentes manières : 1) le geste peut désambiguïser le verbal en apportant un éclairage complémentaire notamment lorsqu'on utilise des mots polysémiques, 2) le geste et le verbal peuvent être redondants, notamment lorsque les mots sont difficiles à comprendre, le geste apporte des indices sur leur signification et 3) le geste peut compléter le verbal notamment lorsque l'enseignant veut éliciter une réponse de la part de l'apprenant tout en donnant un indice sur la réponse attendue via le geste (Tellier, 2016).
- 36 Une deuxième observation concerne le niveau prosodique dans la modalité orale, spécifiquement l'usage des pauses du discours du locuteur expert. Dans le discours à visée pédagogique, l'expert segmente la chaîne parlée pour faciliter la compréhension des mots clés du discours par l'apprenant. Plusieurs études réalisées au LPL ont montré que dans ces pauses du discours, des gestes apparaissaient pour illustrer les mots-clés mis en valeur sémantiquement (Stam & Tellier, 2017 ; Tellier *et al.*, 2013). Par exemple, en classe de langue, une enseignante de FLE explique ce qu'est « la bavette ». Elle dit que c'est un type de viande et que « [c'est des tranches] [(0.33) fines (0.65)] ». Les parties de l'énoncés entre crochets indiquent la production d'un geste et les chiffres entre parenthèses les pauses du discours en secondes. Ici, l'enseignante produit un geste pour illustrer « tranche » (un mot important dans l'explication et potentiellement inconnu) et un autre geste pour « fine ». On remarque que « fine » est encadré de deux pauses qui mettent en valeur ce mot. Le geste iconique qui l'accompagne commence et se termine dans ces pauses (Tellier, 2016). Ce phénomène a été observé à la fois en contexte écologique et semi-contrôlé.
- 37 Une troisième analyse concerne l'adaptation gestuelle. Dans le corpus GTT<sup>12</sup> (« Gesture in Teacher Talk »), de futurs enseignants devaient faire deviner des mots à des partenaires francophones et des apprenants du français. Cette tâche permettait d'évaluer si ces futurs enseignants adaptaient leur gestuelle en fonction du niveau de langue de leur partenaire. Les résultats montrent qu'avec des apprenants du français, ils produisent des gestes plus nombreux, plus longs en durée, plus illustratifs et plus amples (Tellier & Stam, 2012 ; Tellier *et al.*, 2021).

- 38 Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/5167>

## 2.2. Les interactions en ligne

39 Avec la diffusion des ordinateurs et de connexions internet de plus en plus performantes, plusieurs outils de communication médiatisée par ordinateur (CMO) se sont développés et diffusés au fil des années, à partir des premiers échanges de courriel jusqu'à la banalisation récente de la visioconférence et des appels vidéophoniques. Ces nouveaux outils ont, d'une part, fait l'objet d'une projection de la part des interactions de procédés interactionnels provenant des interactions en face à face présentiel, tout en faisant émerger des nouveaux phénomènes communicatifs liés à la multimodalité de différentes formes de CMO (que l'on pense par exemple aux émoticônes dans les clavardages). Certaines recherches au LPL se sont ainsi tournées vers l'émergence de techno-genres discursifs et leur caractérisation par exemple en termes d'ethos. D'autres se sont intéressées au potentiel pédagogique pour la didactique des langues, tant sur le versant de l'acquisition de langues secondes que sur celui de la formation de formateurs.

### 2.2.1. Interactions sur le Web 2.0

40 Plusieurs études au LPL s'appuyant sur le cadre théorique de l'Analyse du Discours Numérique (Paveau, 2017) s'intéressent au fonctionnement des interactions dans les espaces du web 2.0. Elles visent notamment à décrire, caractériser et même typologiser les différents genres numériques émergents dans ces espaces en prenant davantage en compte l'écosystème du web (Paveau, 2013) car la question de la contrainte technologique dans le fonctionnement des genres en ligne est primordiale. Ainsi Combe (2019a) étudie l'énonciation numérique de deux genres prototypiques du web 2.0 que sont le blogue et le vlogue et observe comment le système de gestion de contenu (CMS) de Wordpress et la plateforme YouTube contraignent et inspirent l'écriture numérique. Les différents éléments de la multimodalité numérique (voir 1.3.) sont également étudiés et notamment comment ceux-ci s'articulent les uns avec les autres. On observe ainsi comment le scriptural, le graphisme et la vidéo tissent des liens les uns avec les autres sur un blogue de voyage ou comment des commentaires écrits multiples répondent à une vidéo initiale autocentrée dans le cas d'un vlogue. Toujours à travers l'étude de la mise en discours numérique, Combe (2019b) s'intéresse aussi plus particulièrement à la construction de l'ethos de vlogueurs et met en évidence comment des éléments verbo-oraux, posturo-mimo-gestuels, scripturaux et graphiques au sein d'une vidéo permettent la construction d'un ethos discursif multimodal et intermédial d'un pro-amateur qui se veut compétent, accessible et engageant et de quelle manière ils sont exploités pour tenter de reconfigurer l'éducation à la politique et le débat citoyen entre les internautes (Combe, 2019b). Dans l'exemple ci-dessous, on voit le vlogueur Ludo expliquer dans une vidéo l'affaire Sarkhozy-Kadhafi ; à son discours oral viennent s'ajouter des éléments visuels (comme le livre en fond « Avec les compliments du guide, Sarkhozy-Kadhafi l'histoire secrète »), l'incrustation de la photo de Choukri Ghanem, mais aussi mimo-gestuels puisqu'il mime des guillemets pour accompagner l'appellation argotique « Monsieur Tune numéro 2 » (qui désigne Choukri Ghanem).

Figure 2 : Exemple de multimodalité numérique dans un vlogue



- 41 Enfin, en contexte didactique, on peut observer comment des apprentis tuteurs de français langue étrangère redéfinissent leur ethos numérique en étudiant différents genres discursifs multimodaux comme leur e-portfolio étudiant (Combe, 2020), leur selfie vidéo tutorial (une vidéo de présentation à destination de leurs apprenants de FLE), la rédaction de leur consignes en ligne sur une plateforme asynchrone et enfin au cours de séances de visioconférence que ces tuteurs animent (Combe & Cappellini, 2021).

### 2.2.2. Interactions exolingues

- 42 Deux contextes d'interaction sont le focus de plusieurs études : les interactions télé-tandem et les interactions entre tuteurs et apprenants. Dans les interactions de tandem, deux étudiants de deux langues maternelles différentes et apprenant chacun la langue de l'autre interagissent pour s'entraider dans leur apprentissage. Chaque langue est parlée par les deux interlocuteurs la moitié du temps. Si le tandem se déroule en présentiel, le télé-tandem exploite les logiciels et plateformes de vidéoconférence poste à poste. Une des caractéristiques du télé-tandem est que les apprenants n'ont d'expertise que concernant l'usage de leur langue maternelle, donc sans avoir une expertise (méta)linguistique ni pédagogique. Néanmoins, l'étude de Cappellini (2014) montre que les usagers experts ont recours à la multimodalité de la visioconférence pour étayer l'apprentissage langagier et (inter)culturel des apprenants. Ainsi, par exemple le recours au clavardage apparaît notamment lors des séquences latérales d'étayage langagier, autrement dit des moments où l'interaction est, en quelque sorte, suspendue afin de traiter une panne lexicale ou une incompréhension et ainsi en faciliter l'apprentissage. D'autre part, l'étayage pour l'apprentissage d'éléments culturels s'appuie souvent sur un recours à des navigateurs internet servant à la recherche d'images, de musiques et plus largement d'informations à partager avec

l'interlocuteur. L'exemple suivant montre deux étudiantes dans un télé-tandem franco-chinois aux prises avec une séquence latérale de recherche lexicale, ou séquence potentiellement acquisitionnelle lexicale.

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/5167>

- 43
- 44 Dans cet extrait, on peut noter que dans la phase de sollicitations, l'étudiante chinoise produit des hésitations verbales/pauses remplies (euh) accompagnées par une mimique de *doing thinking* (Liebscher & Reichert, 2012) en déplaçant le regard en haut à droite, pour ensuite opérer une sollicitation avec une alternance codique. L'étudiante française propose une traduction française et, une fois cet élément lexical validé par l'apprenante, elle l'envoie également par le clavier. L'extrait montre ainsi comment les différentes modalités de la visioconférence sont mobilisées pour une production transmodale de la communication articulant plusieurs modes : paraverbal (euh, ah), verbal en deux langues, mimique et écrit.
- 45 Les interactions par visioconférence dans le cadre de l'apprentissage des langues et cultures impliquent des compétences qui ont été définies en termes de littéracies socio-interactionnelles (Ollivier & Puren, 2011), de compétence multimodale (Hauck, 2010) ou encore de compétence techno-sémio-pédagogique (Guichon, 2012). Bien que ces définitions impliquent des ancrages théoriques et éventuellement épistémologiques différents, un certain nombre de traits leur est commun. Le principal de ces traits est la relation qui s'instaure entre un locuteur et un interlocuteur à travers un environnement médiatisé, ce qui implique d'une part une conscience plus ou moins approfondie des fonctionnalités des interfaces et des artefacts techniques, d'autre part une capacité à traduire cette conscience en des plans d'action pendant l'interaction même.
- 46 Les recherches dans le domaine de la didactique des langues et cultures ont souvent associé la théorisation de ces compétences en termes de conscience et de capacité à la nécessité d'une formation pour une réussite de la communication à des fins pédagogiques. Néanmoins, suite à l'étude de Cappellini (2014) force est de constater qu'un certain nombre de sous-compétences sont développées de manière informelle (autrement dit de manière incidente et probablement inconsciente) par la pratique même de ce type d'interaction et d'environnement de communication. En particulier, l'étude de Cappellini et Azaoui (2017) a mis en place une comparaison entre des futurs enseignants de FLE engagés dans un tutorat avec des apprenants dans le cadre d'une formation aux compétences techno-sémio-pédagogiques d'une part, et des étudiants en télé-tandem qui n'ont pas de préparation en ce sens. En s'intéressant aux séquences d'évaluation normative (« SEN », i.e. des séquences latérales où l'expert « corrige » une formulation jugée erronée de l'apprenant), ils ont noté que les tuteurs considérés ne montraient pas forcément un plus grand degré de compétence dans leurs prises de parole. Plus particulièrement, le degré de densité multimodale (i.e. le nombre de modes mobilisés) était similaire entre les deux populations. De plus, bien qu'il y ait un plus grand nombre de SEN chez les tuteurs, le taux de reprises des corrections par les apprenants était majeur en télé-tandem.
- 47 A partir de ce constat, une équipe du LPL a conçu le projet de recherche VAPVISIO (<https://anr.fr/Projet-ANR-18-CE28-0011>), qui vise à déterminer : 1. quelles sont les

compétences techno-sémio-pédagogiques nécessaires à l'enseignement - apprentissage des langues par visioconférence ; 2. lesquelles, parmi ces compétences, sont développées par la simple pratique de ces interactions pédagogiques ; 3. corrélativement, quelles compétences au contraire nécessitent une formation formelle pour être développées. De plus, dans la littérature existante, les compétences techno-sémio-pédagogiques ne sont étudiées que par l'observation d'enregistrements audio-visuels des interactions de visioconférence et/ou par des productions après-coup des apprenants explicitant leurs logiques d'action par des questionnaires ou des entretiens, par exemple d'auto-confrontation. Le projet Vapvisio a élaboré un cadre théorique et méthodologique intégrant l'exploitation de l'oculométrie afin d'élaborer des hypothèses d'analyse sur le processus d'action-perception-interprétation (Van Lier, 2004) pendant l'interaction même. En d'autres mots, en analysant les mouvements oculaires d'un interlocuteur, il est possible d'émettre des hypothèses sur sa connaissance des interfaces graphiques de visioconférence et sur la sélection des éléments pertinents pour accomplir une (inter)action, ainsi que leur mobilisation éventuelle (Cappellini, 2021 ; Cappellini & Hsu, 2022). Cette méthodologie permet ainsi de développer des analyses non seulement sur la mobilisation de plusieurs modes sémiotiques, mais aussi sur la conscience de certains modes et de leurs interactions mutuelles.

### 2.3. Les interactions hybrides

- 48 Les interactions hybrides mêlent face-à-face et distanciel et peuvent avoir lieu selon différentes configurations. Une recherche interdisciplinaire et collective à plusieurs laboratoires, dont le LPL (voir fiche corpus "Présences numériques"<sup>13</sup> dans ce numéro), a étudié un séminaire doctoral poly-artefacté (Develotte *et al.*, 2021), c'est-à-dire un séminaire dont une partie des participants est physiquement en présence et l'autre en présence via différents artefacts.
- 49 Ces artefacts étaient au nombre de trois : 1) une tablette articulée sur un socle (le robot de télé-présence Kubi), 2) un robot à taille humaine mobile sur roulettes (le robot de télé-présence Beam) et 3) une plateforme multimodale interactive (Adobe Connect) qui présentait, entre autres modules, un espace de visioconférence, un espace de clavardage, une prise de notes collective et un partage de documents. Chaque artefact étant lui-même actionné au travers d'écrans multiples (d'ordinateurs, de tablettes, de smartphones, etc.).
- 50 Un tel contexte de communication multimodale poly-artefacté s'avère donc éminemment complexe tant à expérimenter pour les participants, qu'à étudier et retranscrire pour les chercheurs. En effet, une grande partie des éléments de la multimodalité numérique (voir partie 1.3) doit être pris en compte comme 1) le verboral et 2) le posturo-mimo-gestuel des participants en présence mais également par visioconférence et via l'écran des robots de télé-présence ; 3) le scriptural du clavardage, de la prise de notes collective, des documents partagés ; 4) les différents éléments d'audio et de vidéo des robots de télé-présence et de la visioconférence ; 5) la cinétique des robots de télé-présence mais aussi des locuteurs à l'écran et en présence physique, etc.
- 51 En outre, la multimodalité concerne également le dispositif de captation des données puisque des données scripturales (questionnaires écrits), mais aussi des données audio

(via des micros) et vidéo (par caméras externes, captures dynamiques d'écran, etc.) ont été recueillies.

- 52 Enfin, comme nous l'avons mentionné précédemment dans le cadre théorique, l'édition numérique offre désormais la possibilité de publier et d'illustrer de manière multimodale les résultats de la recherche. C'est le choix qui a été fait dans l'édition augmentée de l'ouvrage la « Fabrique de l'interaction parmi les écrans : formes de présences en recherche et en formation » (Develotte *et al.*, 2021). Des capsules vidéo ont été spécialement conçues à partir des données pour illustrer les concepts (Lascar & Dujour, 2021) en retranscrivant de manière multimodale (et parfois à partir de certains procédés cinématographiques) l'analyse des interactions. L'exemple ci-dessous est un gif animé qui rend compte d'un mouvement d'esquive des caméras de captation par un participant-chercheur étudié dans l'article dédié à l'étude de la question de la politesse [voir l'article "Politesse, impolitesse et violence verbale dans les interactions humaines"<sup>14</sup> dans ce numéro] dans ce contexte poly-artefacté.

Figure 3 : Gif animé illustrant dans l'édition augmentée un exemple d'acte de coopération dans l'interaction (*ibid.*)



### 3. Perspectives

- 53 Au terme de cet article qui met en évidence les grandes directions de recherche effectuée au LPL sur la multimodalité, plusieurs perspectives ont été identifiées. A titre d'exemple, étudier le phénomène de défaillance interactionnelle semble particulièrement intéressant notamment pour montrer comment les locuteurs puisent dans leurs ressources multimodales pour surmonter ces défaillances. Des phénomènes différents les uns des autres sont à observer : disfluences, recherches lexicales, incompréhensions et réparations, explications et enfin anticipation de la défaillance à un travers un discours didactique multimodale.
- 54 Une première piste de recherche qui pourrait être explorée est la nature multimodale des disfluences en interaction. Quelques études, dont celles de Dodane *et al.* (2019), Graziano et Gullberg (2018) Kosmala *et al.*, (2019) concourent à établir que la disfluence est un phénomène multimodal et que les indices visuels sont intéressants à prendre en compte lors des études sur la parole spontanée. Par exemple, dans un contexte d'apprentissage de langue seconde, Graziano et Gullberg, (2018) ont mis en évidence une synchronisation forte entre la suspension de la parole et celle du geste. Esposito et Marinaro (2007) ont également montré une relation ente les pauses et les suspensions de la main (*hold*) aussi bien chez l'enfant que chez l'adulte. Les mêmes résultats ont été trouvés lors du bégaiement chez des sujets bègues (McNeil, 2005). Les études de Cook *et al.*, (2009) montrent que lorsque les locuteurs choisissent des formes langagières moins habituelles, leurs gestes sont plus nombreux et le nombre de leurs disfluences augmente. Au contraire, les disfluences analysées par Karpinski et Jarnolowicz-Nowkow (2020) lors de dialogues au cours de tâches comme la dénomination d'image ne présentent pas cette même synchronisation avec le geste. La synchronisation de disfluences et de gestes ferait donc encore débat. La participation des disfluences morphosyntaxiques à l'élaboration de l'énoncé encourage donc à envisager le couplage au moins de certaines disfluences avec des paramètres verbaux et gestuels : observe-t-on seulement une synchronisation de la suspension du geste et de la parole ou existe-t-il des gestes spécifiques liés à la recherche lexicale ? Une perspective de recherche au LPL pourrait être de comparer l'apparition des disfluences dans les discours en L1 vs en langue étrangère pour voir si la maîtrise de la langue affecte la façon dont ce phénomène s'exprime multimodalement. Les éléments à prendre en compte seraient les pauses (vides ou remplies), la morpho-syntaxe, les gestes co-verbaux, et l'orientation du regard. Ce dernier point semble particulièrement pertinent car lors de la disfluence (notamment pour recherche lexicale) le regard peut être dirigé soit vers l'interlocuteur pour solliciter son aide ou soit ailleurs pour se concentrer et couper de son champ de vision l'interlocuteur comme on peut le voir dans la figure ci-dessous.

Figure 4 : Orientation du regard pendant une recherche lexicale multimodale (Stam &amp; Tellier, 2017)



(on dit)

qu'on euh {2268ms}

(one says)

that one hum {2268ms}

- 55 Parmi les perspectives à explorer dans les recherches sur la multimodalité, on peut compter l'étude des alignements conversationnels. Dans le modèle proposé par Pickering et Garrod (2021), l'alignement est un phénomène psychologique concernant les représentations mentales d'interlocuteurs, phénomène qui se reflète dans l'articulation et souvent la reprise de comportements communicatifs. Si l'alignement a initialement été étudié par le phénomène du *priming* à différents niveaux langagiers (lexique, morphosyntaxe...), récemment on a pu constater un intérêt pour la communication dans sa multimodalité, incluant entre autres les gestes et les mimiques faciales. Un premier pas dans cette direction a été proposé par Cappellini, Holt et Hsu (2022), qui se proposent d'étudier comment, dans les interactions exolingues en ligne, l'alignement se concrétise aux niveaux lexicaux, morpho-syntaxique et mimique.
- 56 En incluant l'étude des mimiques faciales, une étude du sourire des interlocuteurs en phase de réparation orale, pourrait permettre de mettre en exergue l'importance de cette expression faciale dans ce procédé interactionnel. Une piste de recherche serait également d'interroger la place du sourire dans la recherche lexicale (cf. extraits présélectionnés).
- 57 Enfin, il semble que l'on pourrait aussi dans les recherches sur la multimodalité numérique étudier à partir de la complexité du schéma interactionnel de la plateforme YouTube (un discours vidéo auquel répondent des commentaires écrits en asynchronie) comment les vlogueurs anticipent la possible défaillance interactionnelle en recourant à un discours didactique (Moirand, 1993) qui sollicite différents éléments multimodaux.

---

## BIBLIOGRAPHIE

Amoyal, M., Priego-Valverde, B., & Rauzy, S. (2020) PACO: A corpus to analyze the impact of common ground in spontaneous face-to-face interaction, *Proceedings of the 12<sup>th</sup> Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, p. 621–626.

Anis, J. (1998) *Texte et Ordinateur. L'écriture réinventée ?*, Bruxelles : De Boeck Université.

- Argyle, M. (1988) *Bodily communication* (2nd ed.), London : Methuen.
- Aubergé, V., & Cathiard, M. (2003) Can we hear the prosody of smile?, *Speech Communication*, 40 (1-2), p. 87-97.
- Barrier, G. (2013) La sémiologie du corps dans les interactions, in G. Barrier, *Les langages du corps en relation d'aide : La communication non verbale au-delà des mots*, Issy-les-Moulineaux : Esf, p. 41-74.
- Bavelas, J., Gerwing, J., & Healing, S. (2014) Hand and facial gestures in conversational interaction, *The Oxford handbook of language and social psychology*, p. 111-130.
- Bavelas, J. B., & Gerwing, J. (2007) Conversational hand gestures and facial displays in face-to-face dialogue, *Social communication*, p. 283-308.
- Beliao, J., Lacheret, A. (2013) Disfluencies and discursive markers. When prosody and syntax plan discourse, *DISS 2013 The 6<sup>th</sup> workshop on disfluency in spontaneous speech August 2013*, Stockholm Sweden, 54 (1), p. 5-9.
- Bertrand, R., Blache, P., Espesser, R. Ferre, G., Meunier, C., Priego-Valverde, B. & Rauzy, S. (2008) Le CID- Corpus of Interactional Data - Annotation et exploitation multimodale de Parole conversationnelle. *Traitement automatique des Langues*, 49-3, p. 105-134.
- Birdwhistell, R. L. (1968) L'analyse kinésique, *Langages* (10), p. 101-106.
- Bolly, C.T., Christodoulides, G. & Simon, A.C., (2017) Disfluences et vieillissement langagier. De la base de données VALIBEL aux corpus outillés en français parlé, *Corpus* [En ligne], 15, 2016, mis en ligne le 15 janvier 2017, consulté le 17 avril 2022. <https://doi.org/10.4000/corpus.3006>
- Cappellini, M. (2014) *Modélisation systémique des étayages dans un environnement de tandem par visioconférence pour le français et le chinois langues étrangères. Une étude interactionniste et écologique du soutien au développement de la compétence de communication*. Thèse soutenue le 13 novembre 2014 à l'Université Lille 3 SHS, laboratoire Savoirs, Textes, Langage (STL, UMR 8163 CNRS).
- Cappellini, M. (2021) Une approche multimodale intégrant l'oculométrie pour l'étude des interactions télécollaboratives par visioconférence, *Les Cahiers de l'ASDIFLE*, 31, p. 99-120.
- Cappellini, M., & Azaoui, B. (2017) Sequences of normative evaluation in different pedagogical settings through desktop videoconference, *Language Learning in Higher Education*, 7(1), p. 55-80.
- Cappellini, M., Holt, B., & Hsu, Y.-Y (2022) Multimodal alignment in telecollaboration: a methodological exploration, *System Journal*. <https://doi.org/10.1016/j.system.2022.102930>
- Cappellini, M., & Hsu, Y.-Y (2022) Multimodality in Webconference-Based Tutoring: An Ecological Approach Integrating Eye-Tracking. *ReCALL Journal*, 34-3, p. 255-273.
- Cole J., Hasegawa-Johnson M., Shih C., Heejin Kim, Eun-Kyung Lee, Hsin-yi Lu, Yoonsook Mo, Tae-Jin Yoon, (2005) Prosodic parallelism as a cue to repetition and error correction disfluency, *Proceedings of DISS'05. Disfluency in spontaneous speech Workshop, 10-12 september 2005, Aix-en-Provence*, p. 1-6.
- Collard, Ph., Corley M. & MacGregor L.J. (2008) Attention orienting effects of hesitations in speech: Evidence from ERPs, *Exp Psychol Learn Mem Cogn*, 34(3), p. 696-702.
- Colletta, J.- M. (2004) Le développement de la parole chez l'enfant âgé de 6 à 11 ans : corps, langage et cognition, vol. 254, Editions Mardaga.
- Combe, C. (2019a) Les genres numériques de la relation, *Langage et Société*, N° 167 (2), p. 51-80.

Combe Celik, C. (2019b) Quand le technogène de discours « vlogue » éduque à la politique : le cas français d'« Osons causer », in George Eric (Ed.) *Numérisation généralisée de la société et enjeux sociopolitiques de la culture et de la communication*, ISTE.

Combe, C. (2020) Former de futurs professionnels du FLE au genre du e-portfolio : Une Étude exploratoire. Enjeux et défis du numérique pour l'enseignement universitaire / Chancen und Herausforderungen der Digitalisierung in der Hochschullehre, Peter Lang, Studien zur Translation und Interkulturellen Kommunikation in der Romania, p. 93-117.

Combe, C. et Cappellini, M. (2021) La construction d'un ethos discursif multimodal de futurs enseignants en ligne, *Le Français dans le monde, Recherches et applications*, p. 33-45.

Cosnier, J., Berrendonner, A., Coulon, J. & Kerbrat-Orecchioni, C. (1982) *Les voies du langage : communications verbales, gestuelles et animales*, Dunod : Paris.

Couper-Kuhlen, E., & Selting, M. (1996) Towards an interactional perspective on prosody and a prosodic perspective on, *Prosody in conversation: Interactional studies*, 11.

Develotte, C., Kern, R. & Lamy, M.N. (2011) *Décrire la conversation en ligne*, Lyon : ENS Editions.

Develotte, C., Paveau, M.-A. (2017) Pratiques discursives et interactionnelles en contexte numérique. Questionnements linguistiques. *Langage et société*, 2(2-3), p. 199-215. <https://doi.org/10.3917/lis.160.0199>

Develotte, C., Bouquain, A., Codreanu, T., Combe, C., Domanchin, M., El Hachani, M., Furnon, D., Grassin, J.-F., Ibnelkaid, S., Remon, J. & Vincent, C. (2021) *Fabrique de l'interaction parmi les écrans : formes de présences en recherche et en formation*, édition augmentée, Montréal : Les Ateliers de Sens Public. <http://ateliers.sens-public.org/fabrique-de-l-interaction-parmi-les-ecrans/index.html>

Dodane, C., Boutet, D., Didirkova, I., Hirsch, F., Oumi, S. & Morgenstern, A. (2019) An integrative platform to capture the orchestration of gesture and speech, *Proceeding of the 6<sup>th</sup> gesture and speech in interaction (GESPIN6)*. <https://doi.org/10.1037/0278-7393.34.3.696>.

Drissi, S. (2011) Apprendre à enseigner par visioconférence : Étude d'interactions pédagogiques entre futurs enseignants et apprenants de FLE, thèse de doctorat, ENS Lyon.

Ekman, P., & Friesen, W. V. (2003) *Unmasking the face: A guide to recognizing emotions from facial clues* (Vol. 10), Ishk.

El Haddad, K., Chakravarthula, S. N., & Kennedy, J. (2019) Smile and laugh dynamics in naturalistic dyadic interactions: Intensity levels, sequences and roles, in *2019 International Conference on Multimodal Interaction*, p. 259-263.

Esposito, A. & Marinaro, M., (2007) What pauses tell us about speech and gesture partnership, in A. Esposito et al. (Eds.) *Fundamentals of verbal and nonverbal communication and the biometric issue*. Amsterdam, Berlin, Oxford, Tokyo, Washington: IOS Press, p. 45-58.

Fagel, S. (2010) Effects of smiling on articulation: Lips, larynx and acoustics, in *Development of multimodal interfaces: active listening and synchrony*, Springer, Berlin, Heidelberg, p. 294-303.

Ferguson, C. A. (1975) Toward a characterization of English foreigner talk, *Anthropological linguistics*, p. 1-14.

Ferré, G. (2011) Analyse multimodale de la parole, *Rééducation Orthophonique*, 246, p. 73-85.

Fraundorf, Scott H. and Watson, Duane G. (2011) The disfluent discourse: Effects of filled pauses on recall. *J Mem Lang*, 2011 Aug 1, 65(2), p. 161-175. <https://doi.org/10.1016/j.jml.2011.03.004>

Fridlund, A. J. (2014) *Human facial expression: An evolutionary view*, Academic Press.

- Goffman, E. (1974) *Les rites d'interaction*, Paris : Editions de Minuit.
- Ghio, A., Tellier, M. (2012) Les mouvements de la parole : de l'articulation à la gestuelle, in *Actualités scientifiques dans les Hautes-Alpes*, 83, p. 5-10.
- Gironzetti, E. (2017) *Multimodal and eye-tracking evidence in the negotiation of Pragmatic intentions in dyadic conversations: The case of humorous discourse*, unpublished dissertation, Texas A&M University-Commerce.
- Gironzetti, E., Attardo, S., & Pickering, L. (2016) Smiling, gaze, and humor in conversation: A pilot study. In L. Ruiz-Gurillo, *Metapragmatics of Humor: Current research trends*.
- Goujon, A. (2019) Indices d'incompréhension et séquences de réparation dans l'interaction en face-à-face : une analyse multimodale (Dissertation doctorale non publiée), Aix-Marseille.
- Graziano, M., Gullberg M. (2018) When speech stops, gesture stops: evidence from developmental and crosslinguistic comparisons, *Frontiers in Psychology*, 9 : 879, <https://doi.org/3389/fpsy.2018.879>
- Guichon, N. (2012) L'apprentissage des langues médiatisé par les technologies (ALMT) – Étude d'un domaine de recherche émergent à travers les publications de *la revue Alsic*. *Alsic*, Vol. 15, n° 3, 2012, mis en ligne le 15 novembre 2012, consulté le 25 mai 2021. URL : <http://journals.openedition.org/alsic/2539>.
- Guichon, N. (2013) Une approche sémio-didactique de l'activité de l'enseignant de langue en ligne : réflexions méthodologiques, *Éducation et didactique* [En ligne], 7-1, 2013, mis en ligne le 31 janvier 2015, consulté le 25 mai 2021. URL : <http://journals.openedition.org/educationdidactique/1679>
- Guichon, N. & Tellier, M. (2017) *Enseigner l'oral en ligne, une approche multimodale*, Paris : Didier.
- Hauck, M. (2010) Telecollaboration: At the interface between multimodal and intercultural communicative competence, in S. Guth & F. Helm (Eds.), *Telecollaboration 2.0: Language, Literacies and Intercultural Learning in the 21<sup>st</sup> Century*, Oxford: Peter Lang, p. 219-244.
- Herring, S.-C. (1996 ed.) *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*, Pragmatics and Beyond series, Amsterdam : John Benjamins.
- Herring, S.-C. (2015) New frontiers in interactive multimodal communication, in A. Georgapoulou & T. Spilloti (Eds.), *The Routledge handbook of language and digital communication*, London : Routledge.
- Herring S.C. and Androutsopoulos J. (2015) Computer-Mediated Discourse 2.0, in *The Handbook of Discourse Analysis*, Tannen, Hamilton Schiffrin John Wiley & Sons, Inc.
- Homke, P., Holler, J., & Levinson, S. C. (2018) Eye blinks are perceived as communicative signals in human face-to-face interaction, *PLoS one*, 13(12).
- Karpinski M. & Jarnolowicz-Nowkow E. (2020) Prosodic and gestural features of phrase-internal disfluencies in spontaneous polish speech
- Kendrick, K. H., & Holler, J. (2017) Gaze direction signals response preference in conversation, *Research on Language and Social Interaction*, 50(1), p. 12-32.
- Kopp, S., Allwood, J., Grammer, K., Ahlsén, E., & Stocksmeier, T. (2008) Modeling embodied feedback with virtual humans, in J. G. Cardonell & J. Siekmann (Eds.), *Modeling communication with robots and virtual humans*, Berlin: Springer, p. 18-37.

- Kosmala L., Candea M. & Morgenstern A. (2019) Synchronization of (Dis)fluent Speech and Gesture: A Multimodal Approach to (Dis)fluency, *Gesture and Speech in Interaction*, 2019, Paderborn, Germany. (hal-02360613)
- Kress G. & Van Leeuwen T. (2001) *Multimodal Discourse*, Oxford University press.
- Lascar J. & Dujour O. (2021) Enjeux techniques et défis méthodologiques de l'ingénierie de terrain au service de la recherche, in Develotte, C. et coll. (2021) <http://ateliers.sens-public.org/fabrique-de-l-interaction-parmi-les-ecrans/annexe.html>.
- Long M. H. (1983) Native speaker/non-native speaker conversation and the negotiation of comprehensible input, *Applied Linguistics*, 4(2), p. 126-141.
- McNeill, D. (1992) *Hand and Mind: What gestures reveal about thought*, Chicago : The University of Chicago Press.
- McNeill D. (2005) *Gesture and thought*, Chicago : The University of Chicago Press.
- Moirand, S. (1993) Autour de la notion de didacticité, *Les Carnets du Cediscor* [En ligne], 1, mis en ligne le 28 août 2009, consulté le 30 septembre 2021. URL : <http://journals.openedition.org/cediscor/600>.
- Mondada L. (2005) La constitution de l'origo déictique comme travail interactionnel des participants : une approche praxéologique de la spatialité, *Intellectica*, 41(2), p. 75-100.
- Pallaud B., Bertrand R. (2020) Espaces interruptifs, interruptions suspensives et disfluentes en français parlé dans les huit dialogues du CID, in Fabrice Hirsch, Ivana Didirkova, Christelle Dodane (Eds.), *Manuel de pausologie. Recueil de recherches sur les pauses présentes dans la parole et le discours*, L'Harmattan, Langue et Parole, p. 111-130.
- Pallaud, B., Bertrand, B., Prevot, L., Blache, P. & Rauzy, S. (2019) Suspensive and Disfluent Self Interruptions in French Language Interactions, in Liesbeth Degand, Gaëtanelle Gilquin, Laurence Meurant, Anne Catherine Simon (Eds.) *Fluency and Disfluency across Languages and Language Varieties. Corpora and Language in Use Proceedings 4, Proceedings 4*, p. 109-138.
- Paveau, M.-A. (2019) Technographismes en ligne. Énonciation matérielle visuelle et iconisation du texte, *Corela*, HS-28, 2019, mis en ligne le 11 septembre 2019, consulté le 25 mai 2021. URL : <http://journals.openedition.org/corela/9185>
- Paveau, M.-A. (2017) *L'analyse du discours numérique. Dictionnaire des formes et des pratiques*, Paris : Hermann.
- Pickering, M., & Garrod, S. (2021) *Understanding dialogue*, Cambridge University Press.
- Porquier R. (1994) Communication exolingue et contextes d'appropriation : Le continuum acquisition/apprentissage, *Bulletin VALS-ASLA*, p. 159-169.
- Priego-Valverde, B., Bigi, B., Attardo, S., Pickering, L. & Gironzetti, E. (2018) Is smiling during humor so obvious? A cross-cultural comparison of smiling behavior in humorous sequences in American English and French interactions, *Intercultural Pragmatics*, 15(4), p. 563-591.
- Priego-Valverde, B., Bigi, B. & Amoyal, M. (2020) "Cheese!": a corpus of face-to-face French interactions. A case study for analyzing smiling and conversational humor, in *Proceedings of The 12th Language Resources and Evaluation Conference*, p. 467-475.
- Py B. (2000) Didactique des langues étrangères et recherche sur l'acquisition. Les conditions d'un dialogue, *Études de Linguistique Appliquée*, 120, p. 395-404.

- Reichert, T., & Liebscher, G. (2012) Positioning the expert: Word searches, expertise, and learning opportunities in peer interaction, *The Modern Language Journal*, 96(4), p. 599-609.
- Roubaud M-N. (2004) Du bon usage des amorces dans la transcription des corpus, *Recherches sur le français parlé*, 18, p. 163-184.
- Sacks, H., Schegloff, E., & Jefferson, G. (1974) A Simplest Systematics for the Organization of Turn-taking for conversation, *Language*, 50 (4), p. 696-735.
- Shriberg E. E. (1999) Phonetic Consequences of Speech Disfluency, in *Proceedings of the International Congress of the Phonetic, ICPHS 99*, p. 619-622.
- Shriberg E.E. (1994) *Preliminaries to a Theory of Speech Disfluencies*, PhD thesis, University of California at Berkeley.
- Shriberg E.E. (1995) Acoustic properties of disfluent repetitions, in *Proc. International Congress of Phonetic Sciences*, 4, Stockholm, Sweden, p. 384-387.
- Stam G., Tellier M. (2017) The sound of silence: The functions of gestures in pauses in native and non-native interaction, in R. B. Church, M. W. Alibali, & S. D. Kelly (Éds.), *Why Gesture? How the hands function in speaking, thinking and communicating* (7), John Benjamins Publishing Company, p. 353-377 <https://doi.org/10.1075/gs.7.17sta>
- Stivers, T. (2008) Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation, *Research on language and social interaction*, 41(1), p. 31-57.
- Tabacaru S., Lemmens M. (2014) Raised eyebrows as gestural triggers in humour: The case of sarcasm and hyper-understanding, *European Journal of Humour Research*, 2 (2), p. 11-31.
- Tartter, V. C. (1980) Happy talk: Perceptual and acoustic effects of smiling on speech, *Perception & psychophysics*, 27(1), p. 24-27.
- Tellier M. (2008) Dire avec des gestes, *Le Français dans le monde, Recherches et applications*, 44, p. 40-50.
- Tellier M. (2016) Prendre son cours à bras le corps. De l'articulation des modalités kinésiques avec la parole, *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle*, 13(13-1), Article 1. <https://doi.org/10.4000/rdlc.474>
- Tellier M., Stam, G. (2012) Stratégies verbales et gestuelles dans l'explication lexicale d'un verbe d'action, In Véronique Rivière, *Spécificités et diversité des interactions didactiques*, Riveneuve, p. 357-374.
- Tellier M., Stam G. & Bigi B. (2013) Gesturing While Pausing In Conversation: Self-oriented Or Partner-oriented? *Proceedings of TIGER- Tilburg Gesture Research Meeting Conference*, 5. <http://tiger.uvt.nl/list-of-accepted-papers.html>
- Tellier M., Stam G. & Ghio, A. (2021) Handling Language: How future language teachers adapt their gestures to their interlocutor, *Gesture*, 20(1), 30-62.
- Theonas, G., Hobbs, D., & Rigas, D. (2008) Employing Virtual Lecturers' Facial Expressions in Virtual Educational Environments, *IJVR*, 7(1), p. 31-44.
- Van Lier, L. (2004) *The Ecology and Semiotics of Language Learning: a Sociocultural Perspective*. Dordrecht: Kluwer Academic Publishers.
- Wachs S., Weber C. (2021) Ecrits connectés, nouvel écosystème, nouvelles normes : questions épistémologiques. *Recherches et applications*, CLE International / Français dans le monde, n° 69.
- Winkin, Y. (1981) *La nouvelle communication*, Paris : Le seuil, 372 p.

Wooffitt, R. (2005) *Conversation analysis and discourse analysis: A comparative and critical introduction*, Sage.

## NOTES

1. <https://journals.openedition.org/tipa/5210>
2. Le groupe de travail IMPEC rattaché au laboratoire ICAR mène depuis plusieurs années des recherches sur les interactions par écrans en partant d'ancrages disciplinaires différents (principalement SDL, sciences cognitives et SIC) En ligne : <https://impec.sciencesconf.org/resource/page/id/23>
3. <https://journals.openedition.org/tipa/5424>
4. <https://journals.openedition.org/tipa/pdf/5525>
5. <https://journals.openedition.org/tipa/pdf/5773>
6. <https://journals.openedition.org/tipa/5424>
7. <https://journals.openedition.org/tipa/5424>
8. <https://journals.openedition.org/tipa/pdf/5773>
9. Le terme d'*auto-interruption* réfère à l'arrêt (momentané), par le locuteur lui-même, de sa production verbale par opposition à l'*hétéro-interruption* où l'énoncé est interrompu par un autre locuteur, un bruit ou un événement.
10. <https://journals.openedition.org/tipa/pdf/5770>
11. <https://journals.openedition.org/tipa/5424>
12. <https://journals.openedition.org/tipa/pdf/5790>
13. <https://journals.openedition.org/tipa/pdf/5795>
14. <https://journals.openedition.org/tipa/5293>

---

## RÉSUMÉS

Dans notre contribution, nous allons aborder la notion de multimodalité dans les interactions humaines. Dans un premier temps, nous mettrons au jour les différentes définitions du concept de multimodalité actuellement utilisées dans les recherches en sciences du langage. Dans un deuxième temps, nous détaillerons certaines recherches actuelles sur la multimodalité menées au LPL en montrant la variété des cadres théoriques et méthodologiques mobilisés ainsi que des situations d'interaction considérées, allant du face-à-face présentiel aux interactions en ligne, de la conversation entre locuteurs d'une même langue à la communication exolingue. Enfin, dans une dernière partie, nous montrerons que cette variété n'empêche pas la complémentarité, ni des ressources langagières mobilisées par les participants, ni des cadres théoriques et méthodologiques qui en rendent compte. Nous concluons ce chapitre en proposant quelques pistes de recherche future notamment à partir du phénomène de défaillance interactionnelle.

### 1. Le concept de multimodalité

La communication interpersonnelle (Goffman, 1974) est « multimodale » dans son acception la plus large, c'est-à-dire en prenant en compte aussi bien les travaux de recherche menés en sciences du langage sur l'oral (Colletta, 2004), sur la gestuelle (Cosnier *et al.* 1982, McNeill, 1992), en analyse conversationnelle (Mondada, 2005) que ceux de la sémiotique sociale (Kress &

Van Leeuwen, 2001) ou encore de la communication médiée par ordinateur (Develotte et al, 2011) et de l'analyse du discours numérique (Paveau, 2017). Dans cette première partie, nous définissons nos différents ancrages théoriques principaux, l'un n'excluant pas l'autre.

### **1.1. Éclairages linguistiques**

Lorsque l'on analyse la parole en interaction en tant que phénomène multimodal, on considère en général trois grandes modalités : le verbal, le vocal (dont la prosodie [Voir l'article "La prosodie au Laboratoire Parole et Langage : histoire, recherches actuelles et perspectives" dans ce numéro]) et les éléments posturo-mimo-gestuels (Colletta, 2004). Une linguistique multimodale prend donc en compte les différentes modalités et les analyses en les mettant en relation. Étudier la prosodie ou analyser le geste (ou d'autres aspects kinésiques tels que les postures, les regards et les mimiques faciales) sans le verbal isolerait les modalités et ne permettrait pas de montrer leur articulation (Ferré, 2011).

### **1.2. Éclairage de la sémiotique sociale**

Une conception de la multimodalité par l'entrée des modalités sensorielles peut être complexifiée si l'on pense qu'une même modalité peut être le canal de plusieurs modes sémiotiques. Ainsi, à travers la modalité visuelle, parfois appelée canal de la vue, il est possible de percevoir plusieurs modes sémiotiques : expressions faciales, gestes et proxémique entre autres. Ainsi, la multimodalité peut être définie comme une caractérisation de l'interaction non seulement en termes de modalités à l'œuvre, mais aussi de *modes sémiotiques*.

Le développement des technologies de l'audio-visuel et du numérique a fait émerger dans les années 1990, une autre vision du concept de multimodalité, ancrée dans le champ de la sémiotique sociale. Les travaux de Kress et van Leeuwen (2001) ont permis de définir de nouveaux contours de cette notion en intégrant les différentes modalités de la communication permise par les différents artefacts utilisés. Ainsi la multimodalité est définie comme le recours massif et conjoint à divers modes d'expression (verbal, visuel, audio, tactile, etc.) dans la communication.

### **1.3. Éclairages émergents dans la « multimodalité numérique »**

Le concept de « multimodalité numérique » (Wachs & Weber, 2021) s'ancre dans plusieurs champs de recherches, proches les uns des autres mais présentant chacun des particularités. Le premier, fondateur et précurseur est le champ de la Communication Médiée par Ordinateur (CMO) qui s'est intéressé, dès le début d'Internet (et même du minitel en France) aux échanges en ligne, principalement sous leur forme textuelle (Herring, 1996 ; Anis, 1998). Viennent ensuite deux champs émergents de la linguistique dans la lignée de la CMO : le champ des Interactions Multimodales Par Ecran qui s'est développé notamment autour des interactions par visioconférence (Develotte, Kern & Lamy, 2011, Develotte & Paveau, 2017 ; Guichon & Tellier, 2017) et celui de l'Analyse du Discours Numérique (Paveau, 2017). Enfin, dans le domaine de la didactique des langues, on soulignera le champ de l'Apprentissage des Langues médiatisé par la Technologie (Guichon, 2012) qui étudie l'intégration des technologies dans les situations de communication spécifiquement liées à l'enseignement-apprentissage des langues.

## **2. Les recherches sur la multimodalité au LPL**

Les recherches sur la multimodalité au LPL se répartissent dans trois grands cadres théoriques (celui de la parole multimodale -verbo-vocal et posturo-mimo-gestuel, celui de la sémiotique sociale et enfin celui de la multimodalité numérique) voire parfois dans un rapprochement des trois. Les chercheur.e.s du LPL qui s'intéressent à la question de la multimodalité étudient chacun des contextes différents comme les interactions en face-à-face, les interactions en ligne et les interactions hybrides.

### **2.1. Les interactions en face-à-face**

Les interactions en face-à-face peuvent être de différentes nature (e.g. conversation, interaction médecin-patient, réunion de travail...). Dans cette section, nous présenterons certaines recherches qui portent sur la conversation en face-à-face. Nos travaux portent principalement

sur la prosodie lors de disfluences (Pallaud *et al.*, 2019) et sur le sourire dans deux phases de la conversation : les transitions thématiques et les phases d'humour (Amoyal *et al.*, 2020). Dans l'étude des interactions en face-à-face, un intérêt est également porté à la communication exolingue (entre participants n'ayant pas la même langue première) (Porquier, 1994). Elles sont particulièrement intéressantes à étudier d'un point de vue multimodal, notamment du point de vue de l'adaptation des modalités pour faciliter l'accès au sens (Tellier *et al.*, 2021).

### **2.2. Les interactions en ligne**

Avec la diffusion des ordinateurs et de connexions internet de plus en plus performantes, plusieurs outils de communication médiatisée par ordinateur (CMO) se sont développés et diffusés au fil des années, à partir des premiers échanges de courriel jusqu'à la banalisation récente de la visioconférence et des appels vidéophoniques. Ces nouveaux outils ont, d'une part, fait l'objet d'une projection de la part des interactions de procédés interactionnels provenant des interactions en face-à-face présentiel, tout en faisant émerger des nouveaux phénomènes communicatifs liés à la multimodalité de différentes formes de CMO (que l'on pense par exemple aux émoticônes dans les clavardages). Certaines recherches au LPL se sont ainsi tournées vers l'émergence de techno-genres discursifs et leur caractérisation par exemple en termes d'ethos. D'autres se sont intéressées au potentiel pédagogique pour la didactique des langues, tant sur le versant de l'acquisition de langues secondes que sur celui de la formation de formateurs.

### **2.3. Les interactions hybrides**

Les interactions hybrides mêlent face-à-face et distanciel et peuvent avoir lieu selon différentes configurations. Une recherche interdisciplinaire et collective à plusieurs laboratoires, dont le LPL (voir fiche corpus Présences numériques), a étudié un séminaire doctoral poly-artefacté (Develotte *et al.*, 2021), c'est-à-dire un séminaire dont une partie des participants est physiquement en présence et l'autre en présence via différents artefacts (une tablette articulée sur un socle, un robot à taille humaine mobile sur roulettes et une plateforme multimodale interactive) et qui présentait, entre autres modules, un espace de visioconférence, un espace de clavardage, une prise de notes collective et un partage de documents. Chaque artefact étant lui-même actionné au travers d'écrans multiples (d'ordinateurs, de tablettes, de smartphones, etc.). Un tel contexte de communication multimodale poly-artefacté s'avère donc éminemment complexe tant à expérimenter pour les participants, qu'à étudier et retranscrire pour les chercheurs.

## **3. Perspectives**

Au terme de ce chapitre qui met en évidence les grandes directions de recherche effectuée au LPL sur la multimodalité, plusieurs perspectives ont été identifiées. A titre d'exemple, étudier le phénomène de « défaillance interactionnelle » semble particulièrement intéressant notamment pour montrer comment les locuteurs puisent dans leurs ressources multimodales pour surmonter ces défaillances. Des phénomènes différents les uns des autres sont à observer : disfluences, recherches lexicales, incompréhensions et réparations, explications et enfin anticipation de la défaillance à travers un discours didactique multimodal.

Une première piste de recherche qui pourrait être explorée est la nature multimodale des disfluences en interaction : observe-t-on seulement une synchronisation de la suspension du geste et de la parole ou existe-t-il des gestes spécifiques liés à la recherche lexicale ? Une perspective de recherche au LPL pourrait être de comparer l'apparition des disfluences dans les discours en L1 vs en langue étrangère pour voir si la maîtrise de la langue affecte la façon dont ce phénomène s'exprime multimodalement.

Parmi les perspectives à explorer dans les recherches sur la multimodalité, on peut compter l'étude des alignements conversationnels. Dans le modèle proposé par Pickering et Garrod (2004, 2021), l'alignement est un phénomène psychologique concernant les représentations mentales d'interlocuteurs, phénomène qui se reflète dans l'articulation et souvent dans la reprise de comportements communicatifs. Si l'alignement a initialement été étudié par le phénomène du

*priming* à différents niveaux langagiers (lexique, morphosyntaxe...), récemment on a pu constater un intérêt pour la communication dans sa multimodalité, incluant entre autres les gestes et les mimiques faciales (Cappellini, Holt et Hsu, 2022).

En incluant l'étude des mimiques faciales, une étude du sourire des interlocuteurs en phase de réparation orale, pourrait permettre de mettre en exergue l'importance de cette expression faciale dans ce procédé interactionnel. Une piste de recherche serait également d'interroger la place du sourire dans la recherche lexicale.

Enfin, il semble que l'on pourrait aussi dans les recherches sur la multimodalité numérique étudier à partir de la complexité du schéma interactionnel de la plateforme YouTube (un discours vidéo auquel répondent des commentaires écrits en asynchronie) comment les vlogueurs anticipent la possible défaillance interactionnelle en recourant à un discours didactique (Moirand, 1993) qui sollicite différents éléments multimodaux.

In our contribution, we will discuss the notion of multimodality in human interactions. First, we will present the different definitions of the concept of multimodality currently used in linguistics. Then, we will focus on some of the current studies on multimodality conducted at the LPL, showing the variety of theoretical and methodological frameworks used as well as the interaction situations considered, ranging from face-to-face to online interactions, from conversations between speakers of the same language to exolingual communication. Finally, in the last part, we will show that despite the variety of approaches, our studies are complementary. We will conclude this chapter by proposing some avenues for future research, particularly based on the phenomenon of interactional failure.

### **1. The concept of multimodality**

Interpersonal communication (Goffman, 1974) is “multimodal” in its broadest sense, if we take into account the research work carried out in linguistics on oral speech (Colletta, 2004), on gestures (Cosnier *et al.* 1982, McNeill, 1992), in conversational analysis (Mondada, 2005) as well as in social semiotics (Kress & Van Leeuwen, 2001) or in computer-mediated communication (Develotte *et al.*, 2011) and digital discourse analysis (Paveau, 2017). In this first part, we define our main theoretical anchors, one not excluding the other.

#### **1.1 Insights from linguistics**

When we analyze speech in interaction as a multimodal phenomenon, we generally consider three main modalities: the verbal, the vocal (including prosody) and the posturo-mimo-gestural elements (Colletta, 2004). A multimodal linguistic perspective thus takes into account the different modalities and analyses them by putting them in relation. Studying prosody or analysing gesture (or other kinesic aspects such as postures, gazes, and facial mimicry) without the verbal would isolate the modalities and would not show their articulation (Ferré, 2011).

#### **1.2 Insights from social semiotics**

A conception of multimodality through the input of sensory modalities can be complexified if we think that a single modality can be the channel for several semiotic modes. Thus, through the visual modality, it is possible to perceive several semiotic modes: facial expressions, gestures and proxemics among others. Therefore, multimodality can be defined as a characterization of interaction not only in terms of the modalities at work, but also in terms of semiotic modes.

In the 1990s, the development of audio-visual and digital technologies led to the emergence of another vision of the concept of multimodality, rooted in the field of social semiotics. The works of Kress and van Leeuwen (2001) have allowed to define new contours of this notion by integrating the different modalities of the communication allowed by the different artifacts. Thus, multimodality is defined as the massive and joint use of various modes of expression (verbal, visual, audio, tactile, etc.) in communication.

#### **1.3 Emerging insights of “digital multimodality”**

The concept of “digital multimodality” (Wachs & Weber, 2021) is rooted in several fields of

research, close to each other but each with its own particularities. The first, founding and precursor field is Computer Mediated Communication (CMC) which has been focused one in online exchanges, mainly in their textual form (Herring, 1996; Anis, 1998), since the beginning of the Internet (and even of the minitel in France). Then come two emerging fields of linguistics in the lineage of CMC: the field of Screen-Based Multimodal Interactions, which has developed in particular around videoconference interactions (Develotte, Kern & Lamy, 2011, Develotte & Paveau, 2017; Guichon & Tellier, 2017) and that of Digital Discourse Analysis (Paveau, 2017). Finally, in the field of language didactics, we will highlight the field of Technology-Enhanced Language Learning (Guichon, 2012) which studies the integration of technologies in communication situations specifically related to language teaching and learning.

## **2. Research on multimodality at LPL**

Research on multimodality at the LPL is divided into three main theoretical frameworks: 1) multimodal speech (verbal, vocal and posturo-mimo-gestural), 2) social semiotics, and finally 3) digital multimodality, and sometimes even a combination of the three. The LPL researchers who are interested in the question of multimodality each study different contexts such as face-to-face interactions, online interactions and hybrid interactions.

### **2.1. Face-to-face interactions**

Face-to-face interactions can be of different nature (e.g. conversation, doctor-patient interaction, work meeting...). In this section, we will present some research on face-to-face conversation. Our work focuses on prosody during disfluencies (Pallaud *et al.*, 2019) and on smiling in two phases of conversation: thematic transitions and humor phases (Amoyal *et al.*, 2020). In the study of face-to-face interactions, there is also interest in exolingual communication (between participants who do not have the same first language) (Porquier, 1994). These are particularly interesting to study from a multimodal perspective, especially from the point of view of the adaptation of modalities to facilitate access to meaning (Tellier *et al.*, 2021).

### **2.2. Online interactions**

With the spread of computers and increasingly powerful Internet connections, several computer-mediated communication (CMC) tools have developed and spread over the years, from the first e-mail exchanges to the recent widespread use of videoconferencing and videophone calls. On the one hand, these new tools have been the object of a projection of interactional processes coming from face-to-face interactions, while on the other hand, new communicative phenomena linked to the multimodality of different forms of CMC have emerged (for example, emoticons in chat rooms). Some research at the LPL has thus turned to the emergence of discursive techno-genres and their characterization, for example in terms of ethos. Others are interested in the pedagogical potential for language didactics, both in terms of second language acquisition and teacher training.

### **2.3. Hybrid interactions**

Hybrid interactions mix face-to-face and distance learning and can take place in different configurations. An interdisciplinary and collective research between several laboratories, including the LPL (see *Présences numériques corpus*), studied a poly-artifact doctoral seminar (Develotte *et al.*, 2021), i.e. a seminar where one part of the participants is physically present and the other part is present via different artifacts (a tablet articulated on a base, a human-sized robot mobile on wheels and an interactive multimodal platform) which featured, among other modules, a videoconference space, a chat space, a collective note-taking and a document sharing. Each artifact was itself operated through multiple screens (of computers, tablets, smartphones, etc.). Such a multimodal poly-artifact communication context is therefore eminently complex to use for the participants, as well as to study and transcribe for the researchers.

## **3. Perspectives**

At the end of this chapter, which highlights the main directions of research carried out at LPL on multimodality, several perspectives are identified. For example, studying the phenomenon of

“interactional failure” seems particularly interesting, especially to show how speakers draw on their multimodal resources to overcome these failures. Different phenomena can be observed: disfluencies, lexical searches, misunderstandings and repairs, explanations and finally anticipation of the failure through a multimodal didactic discourse.

A first avenue of research that could be explored is the multimodal nature of disfluencies in interaction: is there only a synchronization of the suspension of gesture and speech or are there specific gestures linked to lexical research? A research perspective at LPL could be to compare the occurrence of disfluencies in L1 vs. foreign language speech to see if language proficiency affects how this phenomenon is expressed multimodally.

Among the perspectives to be explored in research on multimodality, we can mention the study of conversational alignments. In the model proposed by Pickering and Garrod (2004, 2021), alignment is a psychological phenomenon concerning the mental representations of interlocutors, which is reflected in the articulation and often the repetition of communicative behaviours. While alignment was initially studied through the phenomenon of priming at different language levels (lexicon, morphosyntax, etc.), recently there has been an interest in multimodal communication, including gestures and facial mimicry (Cappellini, Holt and Hsu, 2022).

By including the study of facial mimicry, a study of the smiles of interlocutors in the repair phase could highlight the importance of this facial expression in this interactional process. Another avenue of research would be to question the place of the smile in lexical research.

Finally, it seems that we could also study the complexity of the interactional scheme of the YouTube platform (a video discourse to which written comments respond in asynchrony) and how vloggers anticipate possible interactional failures by resorting to a didactic discourse (Moirand, 1993) which calls on different multimodal elements.

## INDEX

**Mots-clés** : conversationnelle, en face-à-face, en ligne, exolingue, humour, hybride, sourire, web 2.0

**Keywords** : conversational, face-to-face, online, exolingual, humor, hybrid, smile, web 2.0

## AUTEURS

### MARY AMOYAL

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
mary.amoyal@univ-amu.fr

### MARCO CAPPELLINI

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
marco.cappellini@univ-amu.fr

### CHRISTELLE COMBE

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
christelle.combe@univ-amu.fr

**BERTHILLE PALLAUD**

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
berthille.pallaud@orange.fr

**BÉATRICE PRIEGO-VALVERDE**

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
beatrice.priego-valverde@univ-amu.fr

**MARION TELLIER**

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
marion.tellier@univ-amu.fr