



<https://bulletin.vals-asla.ch> · e-ISSN: 2813-7973 · p-ISSN: 1023-2044

Monney, M. (2023). Des voix et des mots : le répertoire sémiotique au prisme du discours rapporté direct. *Bulletin suisse de linguistique appliquée*, 117, 165-184. <https://doi.org/10.5169/vals-1062999>

Cet article est publié sous la licence *Creative Commons Attribution 4.0 International* (CC BY): <https://creativecommons.org/licenses/by/4.0>



© Matthieu Monney, 2023

Des voix et des mots: le répertoire sémiotique au prisme du discours rapporté direct

Matthieu MONNEY

Université de Fribourg

Bureau 3.121

Avenue de Beauregard 13, 1700 Fribourg, Suisse

matthieu.monney@unifr.ch

ORCID: 0009-0006-8493-8005

In recent years, multimodality has played an increasingly important role in language sciences, for example through the notions of 'semiotic repertoires' and 'affordances'. This applies particularly well to the study of direct reported speech (DRS), the staging of which regularly gives rise to strong dramatization. We will examine the specific contribution of the voicing of a DRS, in terms of its affordances and the various constitutive elements of a represented utterance. We will try to see how the voice serves to stage a scene, contributing to the relevant attribution of a DRS, constructing personas and attributing attitudes to them, while at the same time conveying the representative speaker's evaluation of the speaker and/or the utterance represented.

Mots-clés:

discours rapporté, dépicition, répertoire sémiotique, voix, polyphonie.

Keywords:

reported speech, depiction, semiotic repertoire, voice, polyphony.

Tant la mise en forme que l'enchâssement du discours d'autrui [...] expriment un acte unique de relations dialogiques avec ce discours, qui détermine tout le caractère de la transmission, et toutes les transformations de sens et d'accent qui se passent en lui au cours de cette transmission. (Bakhtine 1978: 159)

1. Introduction

Que ce soit comme "little drama" (Tannen 1989), "reenactment" (Sidnell 2006), "changement de peau" (Caillat 2012), "mime" (de Cornulier 1978) ou "depiction" (Clark 2016), le discours rapporté (DR) est assez largement conceptualisé en des termes théâtraux. Cela laisse penser que ce phénomène linguistique, dont l'ubiquité dans les interactions quotidiennes a été relevée depuis longtemps (Vološinov [1929] 2010; Goffman 1981; Authier-Revuz 1992; etc.), engage bien davantage qu'une imbrication syntaxique entre deux phrases¹. Le DR s'avère en fait un terrain privilégié pour articuler des questionnements formels à une perspective davantage pragmatique voire sociologique sur le langage. Vološinov l'annonçait déjà lorsqu'il affirmait vouloir, au moyen de l'étude de "la parole d'autrui", "reconstruire à partir de ses formes la façon dont, à telle ou telle

¹ À laquelle l'étude du DR est parfois réduite, notamment dans les manuels scolaires.

époque de son évolution, la langue perçoit le Mot d'autrui et la parole parlante" ([1929] 2010: 461). Cette idée que le DR "is relevant for the study of language, since it shows how the dialogic nature of language can manifest in language structure" constitue d'ailleurs pour Spronck (2022: 1) la thèse centrale de Vološinov. Le deuxième aspect vers lequel pointe la prééminence de la métaphore dramatique dans l'étude du DR est la centralité de la multimodalité. En effet, si l'on y voit un "changement de peau", c'est bien parce que le DR se présente souvent comme "perceptiblement joué" (Caillat 2012: 116) par le locuteur ou la locutrice. Ainsi, cet objet présente un intérêt évident pour l'étude des liens entre grammaire et multimodalité. Et étant donné que c'est spécifiquement le discours rapporté direct (DD) qui est le plus investi par la théâtralisation, c'est sur lui que notre étude se concentrera.

La multimodalité dans le DD a déjà été bien investiguée, en particulier ces dernières années (Sidnell 2006; Park 2009; Fox & Robles 2010; Caillat 2012; Bavelas et al. 2014; Hodge & Johnston 2014; Stec et al. 2015; Hodge & Cormier 2019; Hsu et al. 2021, etc.). Mais nous voudrions apporter notre pierre à l'ouvrage en reconsidérant cet objet à travers le prisme de certaines réflexions de l'école de Palo Alto sur la communication (Watzlawick et al. 1972), ainsi que des notions de "semiotic repertoire" (Kusters et al. 2017) et d'"affordance" (Gibson 1979; Mahrer 2019 pour une application à la linguistique). Cela permettra d'affiner la contribution spécifique de composantes multimodales dans l'élaboration pluridimensionnelle d'un DD. Ces développements seront exemplifiés au moyen de conversations provenant du corpus oral de français de Suisse romande (OFROM²).

Nous commencerons par une réflexion sémiotique sur l'articulation entre la multiplicité des ressources communicationnelles des sujets parlants (leur répertoire sémiotique) – en particulier la mise en voix (vocalisation³) – et la transmission de types distincts d'informations. Nous définirons ensuite notre objet linguistique, le DD, en relevant ses aspects constitutifs au moyen de propositions de Thompson (1996), avant de présenter sa modélisation en termes de "demonstration" (Clark & Gerrig 1990) ou "depiction" (Clark 2016), qui se prête particulièrement bien à la prise en compte de la multimodalité. Cela nous permettra de creuser les potentialités et modalités de sémantisation offertes par la mise en voix d'un DD.

² <https://ofrom.unine.ch/>. Voir Avanzi et al. (2023).

³ Nous entendons par là l'actualisation vocale d'un énoncé, qu'on peut décrire par la prosodie (intensité, rythme, hauteur intonative) et la qualité de voix, conçue avec Günthner comme subsumant toutes les "paralinguistic cues which a speaker may temporarily use in order to produce a whispery, breathy, falsetto, aspirated voice, etc." (1999: 686). La vocalisation sera mise en contraste avec la mise en mots, c'est-à-dire le choix de formulations.

2. Multimodalité et sémiotique

L'idée de relier différentes ressources à l'expression de contenus distincts n'est certainement pas une idée récente. Mais depuis la seconde moitié du XX^{ème} siècle, elle a été formulée et développée de façon à renouveler profondément la conception du langage et de la communication.

2.1 Contenu/relation et affordances

Dans les années cinquante du siècle passé, aux États-Unis, se cristallisent divers courants de recherche qui repensent fondamentalement la communication, en critiquant le modèle "télégraphique" inspiré de l'ingénierie et de la théorie de l'information, pour développer une conception – que Winkin ([1981] 2000) appellera "orchestrale" – de la communication plus dynamique, interactionnelle, synthétique et attentive à la plurisémiotité.

Cette "nouvelle communication" (Winkin [1981] 2000) cherche entre autres à dépasser une conception trop verbo-centriste de la communication. Celle-ci est envisagée au contraire comme "un composé fluide et polyphonique de nombreux modes de comportement: verbal, tonal, postural, contextuel, etc." (Watzlawick et al. 1972: 47), qui interagissent de façons diverses et complexes. De ce fait, tout comportement devient signifiant, selon le fameux axiome "on ne peut pas ne pas communiquer" (*id.*: 48). Une telle conception favorise évidemment l'inclusion des différentes ressources communicationnelles et la considération de leur interrelation complexe.

L'école de Palo Alto (une autre expression pour désigner à peu près la même constellation de chercheurs·euses) va remettre en question un autre présupposé: la réduction de la communication à une transmission d'information. Or pour Watzlawick et collaborateurs, la communication comporte deux dimensions distinctes: le *contenu*, qui "peut avoir pour objet tout ce qui est communicable" (*id.*: 49); et la *relation*, qui désigne "la manière dont on doit entendre le message, et donc en fin de compte la *relation* entre les partenaires"⁴ (*ibid.*). Cette seconde composante met en jeu des assertions telles que "C'est ainsi que je me vois... C'est ainsi que je vous vois... C'est ainsi que je vous vois me voir" (*ibid.*), et relève de la "métacommunication" (*id.*: 51), d'une "*information sur l'information*" (*id.*: 50).

Dans le cadre de la "nouvelle communication", ce couple *contenu/relation* est relié à une distinction entre deux modes de communication, *digital* et *analogique*. Le premier consiste à désigner des objets par leur nom, le second à les représenter au moyen d'un signe qui leur ressemble (ce qui reprend la classique distinction entre *symbole* et *icône*, développée par exemple chez Peirce 1978). Ils se caractérisent d'ailleurs par des propriétés sémiotiques

⁴ Toutes les citations dans cet article sont reproduites telles qu'elles apparaissent dans le texte original, ce qui vaut également pour le formatage, et donc pour la mise en gras et en italique.

particulières: "le langage **digital** possède une syntaxe logique très complexe et très commode, mais manque d'une sémantique appropriée à la relation. Par contre, le langage **analogique** possède bien la sémantique, mais non la syntaxe appropriée à une définition non-équivoque de la nature des relations" (Watzlawick et al. 1972: 65). Ce qui fait qu'ils tendent à se compléter, le digital se rapportant préférentiellement au contenu, et l'analogique à la relation. Se dessine ainsi nettement l'idée d'affinités entre des modes de communication et des types de signifiés, sur la base d'une analyse précise des caractéristiques sémiotiques du médium. Or cette conception s'avère très productive dans l'analyse d'un objet comme le DD, comme nous allons le voir. Dans les exemples analysés, la mise en gras relève les quotatifs⁵, et le soulignement les discours représentés eux-mêmes. Voici un premier cas⁶:

(1)

1 L1: mais genre vous allez vous donner un peu rdv (.) bientôt
 2 L2: ouais il est **il a même dit** on va boire un verre euh (0.5) **je lui ai dit** ouais mais après mes examens (0.4) **pis il a dit** ouais t'inquiète je comprends % mais euh il est pas très beau

(unifr16a09m)

Ici, l'examen des informations verbales permet de reconstruire un échange liminaire, fait d'une proposition, d'une acceptation différée et d'une réaction adoucissante ou rassurante. Mais la prise en compte de la prosodie laisse entendre bien davantage: en effet, au moyen de l'actualisation vocale du premier DD (avec une montée intonative, un changement de ton et un aspect traînant dans la prononciation), L2 attribue au locuteur représenté une attitude pleurnicheuse, évoquant de façon caricaturale un enfant qui ferait une requête affectée envers un adulte. De cette façon, la locutrice (re)configure un "rapport de places" (Vion 1995) particulier, un positionnement intersubjectif où elle se présente en position haute, corrélativement au rabaissement évaluatif du locuteur représenté et/ou de son énonciation. Ce processus se rapporte bien à la *relation* telle que présentée plus haut, et il émane de façon privilégiée de la prosodie, et par là du mode de communication *analogique*.

Les affinités entre médium ou canal sémiotique et type d'information communiquée se laissent assez bien décrire en termes d'"affordances" (Gibson 1979). Ce concept a été développé d'abord dans l'analyse de la perception, pour envisager en quoi les propriétés physiques des objets induisent chez le sujet percevant une certaine utilisation: la forme du marteau par exemple prédispose à le saisir par le manche pour taper avec la tête (du marteau) sur un objet, tout comme la forme de la chaise invite à s'asseoir sur le siège et à reposer le dos contre le dossier. Cette idée a ensuite été appliquée à de

⁵ C'est-à-dire d'éléments lexicaux faisant office d'introducteurs ou d'"encadreurs" de DR.

⁶ Étant donné l'importance de la prosodie dans cette étude, nous encourageons les lecteurs·trices à aller écouter les exemples sur OFROM, librement consultable en ligne.

nombreux domaines, dont les sciences du langage. Mahrer (2019) par exemple convoque cette notion pour approfondir le modèle de Koch & Oesterreicher (2001), en investiguant les raisons des "affinités" relevées par les romanistes entre code graphique et distance communicative, et code phonique et immédiat communicatif. Plus généralement, il cherche à comprendre "pourquoi le média fonctionne comme un attracteur pour certains appareils formels et pour certains comportements cognitifs et sociaux" (Mahrer 2019: §32). Pour lui, cela doit passer par une analyse "biotechnologique", qui "questionne la manière dont la matérialité des signes conditionne leur utilisation" (*id.*: §34). Cette perspective se retrouve dans les "répertoires sémiotiques", que nous allons brièvement présenter maintenant.

2.2 Les "répertoires sémiotiques"

Le concept de "répertoire sémiotique" a été développé au croisement de recherches sur le langage des signes, le plurilinguisme et la multimodalité. Un *répertoire* dénote l'ensemble des moyens communicationnels des sujets parlants, et invite à considérer "the individual's dynamic use of their linguistic resources in different contexts for meaning-making without regard for socio-cultural boundaries in named languages"⁷ (Kusters et al. 2017: 220). Ce qui revient – formulé autrement – à se décentrer du concept de "language" pour aller vers celui de "linguaging" (Becker 1991). Au lieu d'un système expliquant le rapport entre son et sens (la tâche fondamentale attribuée à la linguistique au moins depuis Saussure), cette perspective envisage la communication comme une activité continue de "sense-making", foncièrement multimodale – investissant une multiplicité de canaux sémiotiques qui s'associent et interagissent de manière complexe –, et définitoire de l'être-au-monde des humains.

Ce n'est pas seulement l'association et la reconfiguration permanente de signes provenant de divers codes, adaptés à des besoins communicationnels et des situations, qui est prise en compte. Le *répertoire* favorise l'analyse des façons dont les ressources, qui se déploient généralement de façon simultanée, interagissent en construisant et structurant l'activité communicative, qui se déploie en phases consécutives. Ces moyens communicationnels tendent par ailleurs à s'organiser en constellations, redéfinies pour chaque situation, et en même temps donnant lieu à des configurations typiques, plus ou moins figées. Un dernier aspect, qui nous intéressera moins pour cette étude spécifiquement mais qui est aussi examiné dans le cadre des recherches autour de ces *répertoires*, est l'accès diversifié des sujets aux ressources communicationnelles ainsi que l'organisation et la stratification de ces dernières dans la

⁷ Considérant notre conception multimodale de la communication, cette affirmation serait plus probante en remplaçant "linguistic resources" par "communicative resources".

société, un questionnement particulièrement aigu dans le cas par exemple de personnes malentendantes.

En résumé, les *répertoires* favorisent l'articulation des "ideologies, histories, potential and constraints" (Kusters et al. 2017: 229). Et pour cette recherche, elle ouvre à l'analyse détaillée de l'interaction entre différents canaux sémiotiques – en particulier ici la vocalisation – et les affordances entre ces derniers et les besoins communicationnels des sujets parlants.

3. Modélisation du DD

Nous allons dans cette section relever les éléments constitutifs d'un DD et présenter le modèle de la "depiction" ou "demonstration" de Clark & Gerrig (1990) et Clark (2016). Mais commençons par délimiter brièvement notre objet. Le DD est un type de DR, acte communicationnel complexe s'il en est, que nous définirons avec Rosier comme cette "mise en rapport de discours dont l'un crée un espace énonciatif particulier tandis que l'autre est mis à distance et attribué à une autre source, de façon univoque ou non" (2008: 20). Cette articulation de deux énonciations relève ainsi d'une "*parole dans la parole, [de] l'énoncé dans l'énoncé*, mais en même temps [...] *[d']une parole sur une parole, [d']un énoncé sur un énoncé*" (Vološinov [1929] 2010: 363). Le DR tient donc à l'intégration hiérarchisée d'une énonciation représentée dans une énonciation représentante, intégration qui est également une évaluation du locuteur représentant à l'égard de l'énonciation qu'il (re)construit. Avec Spronck & Nikitina (2019), cette paire représentant/représenté sera désignée par M pour "matrix" et R pour "represented".

Ce qui distingue le DD parmi les différents types de rapport ou représentation de discours est essentiellement le mode sémiotique par lequel s'opère cette représentation de discours, à savoir l'autonymie ou la monstration: les signes à l'intérieur de R dans un DD sont des autonymes, ce sont des signes renvoyant à des signes. Plus précisément, ils renvoient à des signes en exercice, à des signes en tant qu'ils sont actualisés dans une situation particulière⁸. Sur le plan énonciatif, R construit sa propre énonciation et instaure ainsi un "être-de-discours" (Ducrot 1984) qui configure et se trouve configuré par un ensemble de paramètres ou de fonctions qui lui sont attribuables: cette instance constitue l'*origo* d'un ancrage référentiel, est responsable de la mise en mots et en voix, valide les subjectivèmes et se présente comme origine des points de vue et responsable des actes illocutoires effectués dans le cadre de R. Évidemment, selon les DD, l'accent est mis sur un aspect ou l'autre. On trouve par exemple de nombreuses occurrences où la vocalisation n'est pas investie de manière distinctive, c'est-à-dire où le locuteur représentant n'attribue pas de voix

⁸ Au contraire par exemple des autonymes renvoyant à des types de la langue, comme dans ""Vouloir" régit une subordonnée au subjonctif", emprunté à Authier-Revuz (2020: 11).

spécifique à l'instance dont il représente l'énonciation. C'est d'ailleurs le cas pour le dernier DD de l'exemple (1), que la locutrice présente sans changement de voix notable, au contraire du premier DD analysé plus haut. Mais dans tous les cas, le DD s'interprète sur (au minimum) deux centres énonciatifs: celui de M, s'inscrivant typiquement dans l'énonciation en cours, et celui de R, renvoyant à l'énonciation représentée.

Cette définition du DD reste assez élémentaire, étant donné qu'elle sert avant tout à délimiter un objet en vue d'en considérer certaines propriétés en discours. Elle suffira dès lors pour les besoins de cette étude.

3.1 La composition d'un DD

Le DD peut s'analyser en plusieurs constituants définitoires. Pour ce faire, nous nous appuyons sur le modèle de Thompson (1996), qui a élaboré un cadre global pour l'étude des "signaled voices", c'est-à-dire "any stretch of language where the speaker or writer signals in some way that another voice is entering the text" (*id.*: 506). Selon lui, tout locuteur voulant faire entendre une parole autre doit opérer un certain nombre de choix à l'égard de quatre dimensions qui s'entrecroisent tout en étant relativement indépendantes:

- a) La *voix* ("voice"): celui ou ce qui est présenté comme source de R.
- b) Le *message* ("message"): la façon dont le contenu ou la fonction de R est présenté.
- c) Le *signal* ("signal"): la façon dont le locuteur représentant indique qu'il y a représentation de langage et les relations logiques et syntaxiques entre R et M.
- d) L'*attitude* ("attitude"): l'évaluation par le locuteur rapportant de R ou du locuteur représenté.

Pour chacune de ces dimensions, le locuteur dispose d'un certain nombre d'options, qui peuvent se traduire de diverses façons dans l'élaboration de l'énonciation⁹.

L'intérêt du modèle de Thompson est qu'il définit relativement bien ces différentes composantes, et ce sans engagement trop évident pour quelque paradigme théorique que ce soit, en sciences du langage de manière générale ou dans l'étude du DR spécifiquement. Par ailleurs, les aspects qu'il relève se retrouvent également dans d'autres études du DR ou plus largement du "feuilletage énonciatif" (Détrie 2008). Ainsi, ils correspondent quasiment point

⁹ Nous définissons celle-ci avec le Groupe de Fribourg comme "une action communicative composite, une conduite signifiante globale dans laquelle peuvent entrer, outre l'actualisation d'une clause linguistique, l'exhibition d'indices prosodiques, l'accomplissement éventuel de gestes co-verbaux, et/ou la mise en œuvre simultanée de codes annexes de toutes sortes" (2012: 29).

par point aux fonctions que Günthner reconnaît à la prosodie dans le DR. La mise en voix peut servir pour elle:

- (i) to contextualize whether an utterance is anchored in the reporting world or the storyworld; (ii) to animate the quoted characters and to differentiate between the quoted characters; (iii) to signal the speech activities and the affective stance of the reported characters; (iv) to comment on the reported speech as well as on the quoted characters. (Günthner 1999: 704)

Où (i) correspond au *signal*, (ii) à l'*attribution*, (iii) au *message*, en spécifiant la contribution spécifique de la prosodie, et (iv) à l'*attitude*.

Dans tous les cas, l'élaboration de ces aspects laisse entrevoir les multiples dimensions que l'analyse du DR engage: syntaxico-sémantique, sémiotique, modale, énonciative, argumentative ou encore pragmatique, si l'on veut en donner une description la plus fine possible. Pour cette étude, nous prendrons une à une les composantes présentées plus haut, pour réfléchir à la contribution spécifique de la mise en voix dans leur actualisation. Mais avant cela, nous voudrions préciser de quelle façon peut se concevoir la sémiotique particulière du DD, autrement que par l'autonymie, concept dont l'ancrage logico-sémiotique peut le rendre assez difficile à saisir.

3.2 Sémiotique mimique, dépicition, reenactment

Dans une conception verbo-centriste de la communication, il est assez aisé de faire du DD un problème essentiellement syntaxique, de l'ordre de l'intégration d'une phrase dans une autre phrase (une forme de copier-coller). Mais dès lors que l'aspect foncièrement multimodal de la communication est reconnu, les questions se posent différemment.

C'est précisément là qu'on peut faire intervenir le modèle de la "demonstration" ou "depiction" de Clark & Gerrig (1990) et Clark (2016). Dans cette conception, fortement inspirée de la sémiotique peircienne, trois méthodes basiques composent la communication:

- a) La "description": elle consiste à décrire verbalement un référent selon un système symbolique articulé.
- b) L'"indication": elle fonctionne par pointage de quelque chose, et consiste à le localiser dans l'espace-temps pour un destinataire.
- c) La "demonstration" (Clark & Gerrig 1990) ou "depiction" (Clark 2016): elle consiste à créer physiquement une scène pour en représenter une autre par analogie, "enabling others to experience what it is like to perceive the things depicted" (Clark & Gerrig 1990: 765).

Nous nous concentrerons sur cette dernière, étant donné que le DR est conçu comme un type de dépicition, prototypiquement une "demonstration of what a person did in saying something" (*id.*: 769). Ainsi, loin de se limiter à un copier-coller de mots, le DR se présente comme une spécialisation dans les langues naturelles d'un moyen basique de communication traversant le répertoire sémiotique des sujets parlants. Et Clark propose d'inscrire la dépicition dans une

"staging theory", dans laquelle "[t]o perform a depiction is to stage a scene for recipients to use in imagining the scene depicted" (2016: 329). Cette représentation va activer à la fois un imaginaire conceptuel, dans lequel "we imagine that certain propositions are true" (*ibid.*); et un imaginaire perceptuel, dans lequel "we imagine seeing, hearing, touching, or feeling things" (*ibid.*), sur la base d'un stimulus.

De fait, ce modèle nous semble particulièrement productif pour repenser la définition du DD et favoriser son appréhension multimodale, *i.e.* l'interaction des différents modes de communication dans sa mise en scène. Comme annoncé, c'est spécifiquement la contribution de la vocalisation à la construction pluridimensionnelle du DD qui nous intéressera.

4. Mise en voix et en mots du DD

Nous allons désormais examiner la sémiotisation vocale du *signal*, de la *voix*, du *message* et de l'*attitude*, à l'aide d'enregistrements d'interactions orales, pour mettre au jour quelques configurations possibles. Nous commencerons par le *signal*, étant donné qu'avant d'attribuer une représentation de parole (la *voix*) ou d'envisager son contenu (le *message*), il faut déjà reconnaître qu'il y a un DR, afin de pouvoir répondre à la question: "*Is this the current speaker's 'voice' or someone else's?*"¹⁰ (Couper-Kuhlen 1998: 4). Toutefois, au vu de notre questionnement de départ portant sur les affinités entre les canaux sémiotiques et diverses composantes du "sense-making", l'accent sera mis davantage sur le *message* et l'*attitude*, et même l'*attribution* nous intéressera essentiellement en tant qu'elle peut contribuer à ces deux aspects.

4.1 Le signal

Fonctionnellement, toute marque ou indice¹¹ peut faire office de *signal*, dès lors qu'il permet d'identifier un dédoublement énonciatif, la mise en place d'une dépicition, l'irruption d'une autre instance dans le discours (dont les paramètres ont été listés plus haut). Ce repérage peut s'appuyer sur trois logiques distinctes¹².

La première relève de la distinction ou divergence inter-énonciative, selon laquelle plusieurs composantes d'une énonciation dénotent ou induisent des espaces énonciatifs distincts, signalant de "l'autre" dans le "même" de l'énonciation en cours, y introduisant un "corps étranger" (Authier-Revuz 1978: 51). C'est sous l'expression de "discordanciels de l'énonciation" que Rosier

¹⁰ Cette question est la première des trois que pose selon Couper-Kuhlen l'interprétation d'un DR. Les deux autres sont "*Whose 'other voice' is this?*" (1998: 8), qui concerne la *voix* ou l'*attribution*, et "*How is this 'other voice' being done?*" ou "*What is the speaker doing with this 'other voice'?*" (*id.*: 9), relevant plutôt du *message* et de l'*attitude*.

¹¹ Sur la distinction entre les deux, voir Authier-Revuz (2020: 164 et 429 *ssq.*).

¹² Nous nous appuyons pour leur identification et description à nouveau sur Authier-Revuz (2020).

désigne de son côté "tous les mots, expressions ou constructions qui permettent de produire un décrochage énonciatif et donc de repérer au moins deux espaces énonciatifs, l'un citant, l'autre cité, dans un texte" (2008: 78). Sont candidats par exemple les connecteurs, les ruptures modales ou lexicologiques ou encore les marqueurs graphiques (on pense bien sûr au fameux "deux points ouvrez les guillemets!"). La vocalisation s'avère elle aussi très efficace dans cette logique, étant donné que la voix d'une certaine façon "personnalise" (Mahrer 2014) le signal oral, vu qu'elle est physiquement reliée à un corps émetteur, et donc qu'un changement sensible de voix manifeste une forme de travestissement et construit une *persona* distincte du locuteur de l'énonciation en cours. L'expression de "changement de peau" employée par Caillat (2012) dans l'étude du DD prend tout son sens. Ce type de modulation vocale engage un effort physique particulier, qui débouche occasionnellement sur des ratés de mise en voix, quand le locuteur peine à effectuer ou maintenir ce travestissement.

La seconde logique est de l'ordre de la cohérence intra-énonciative, qui fait tenir ensemble certains segments de l'énonciation, en tant que rapportés au même espace énonciatif, au même centre de référence personnel, spatio-temporel, cognitif, épistémique ou encore corporel.

Enfin, la troisième logique relève de la convergence interdiscursive, selon laquelle le destinataire est capable de reconnaître ou d'identifier la source dont provient ce "corps étranger", source qui peut être un locuteur ou une locutrice bien défini·e dont la manière de parler ou l'intonation par exemple est reconnaissable, autant qu'une formation sociale ou discursive aux contours flous, mais à laquelle les sujets parlants sont tout de même capables d'attribuer une certaine qualité vocale, et donc une forme de corporalité¹³. On peut même aller plus loin, en faisant l'hypothèse que le processus d'attribution de voix peut contribuer à la cristallisation des formations discursives dans l'espace social.

S'il est possible de distinguer théoriquement ces trois dynamiques, elles fonctionnent généralement de concert dans la communication, et c'est surtout par leur mise en relation, leur fonctionnement en "faisceau" (Caillat 2012: 116) que les marques et indices fonctionnent. Par ailleurs, la mise à distance d'un espace énonciatif ouvre naturellement vers l'identification de sa source et donc vers l'*attribution*.

4.2 L'*attribution*

Prototypiquement, l'identification de la source de R se fait par sa verbalisation en tant que sujet du verbe de parole introduisant la séquence représentée. Ce cas de figure ne recouvre toutefois pas toutes les façons de répondre à cette

¹³ Sur le rapport entre l'acte d'énonciation, la mise en place d'une image du locuteur et la corporalité, voir les travaux de Maingueneau sur l'*ethos* (*inter alia* 2002, 2012, 2020).

question pas si évidente, qui figure même dans le titre d'un ouvrage sur le DR: "Who's speaking, please?" (Hanote & Chuquet 2004).

Du côté de M, l'établissement d'une source énonciative par la mise en mots peut passer par toutes les stratégies référentielles imaginables, du nom propre aux syntagmes nominaux en passant par l'anaphore et la deixis. Dans tous les cas, le discours enregistre dans la "mémoire discursive"¹⁴ un certain nombre d'entités, et l'*attribution* consiste à sélectionner la source pertinente sur la base d'instructions sémantico-pragmatiques. Mais l'*attribution* peut également rester dans le vague ou la sous-détermination, typiquement par le biais d'un sujet en "on", dont la plasticité référentielle est reconnue depuis longtemps (voir Atlani 1984; Fløttum et al. 2007; Jacquin 2017 *inter alia*), ou des différentes diathèses (impersonnelle, passive, pronominale typiquement) qui court-circuitent la mention du sujet de parole.

Toujours du côté de la verbalisation, l'*attribution* est possible au moyen d'éléments de R, mais sur la base de mécanismes bien différents du décodage linguistique appuyé par une compétence pragmatique, comme nous l'avons présenté plus haut. Au sein de R, la source est inférable de tous les éléments de mise en mots susceptibles de caractériser distinctivement une entité communicante, sachant qu'*in fine*, "c'est, en fait, tout système de nomination du réel, en tant que parcouru par des divergences d'analyse, d'appréciation, d'évaluation et par des variétés de langue – diachroniques, régionales, sociales, etc. – qui est caractéristique d'un acte d'énonciation donné" (Authier-Revuz 1992: 13). Autrement dit, "[d]ès qu'il y a énonciation, quelque chose de l'ordre de l'ethos se trouve libéré" (Maingueneau 2002: 66). Cette identité langagière se construit par "le regroupement des faits en symptômes, opération de type diagnostic, qui mobilise des ressources cognitives de l'ordre de l'empathie" (*id.*: 56). De ce point de vue, c'est donc autant la verbalisation que la vocalisation qui contribuent à l'*attribution*. Dans les deux cas, certaines composantes de l'énonciation, où qu'elles se situent dans le répertoire sémiotique, sont perçues comme des indices permettant d'élaborer ou d'identifier une source.

Toutefois, la vocalisation a ceci de particulier qu'elle est liée physiquement à son émetteur, qu'elle dépend d'un corps particulier, ce qui en fait un "signal personnalisé (singularisé par sa substance)" (Mahrer 2014: 33). Elle s'apparente ainsi à un indice peircien, défini comme "un signe qui renvoie à l'objet qu'il dénote parce qu'il est réellement affecté par cet objet" (Peirce 1978: 164). Et c'est bien pour cela que la vocalisation de R dans un DD, dès lors qu'elle fait partie des "depictive aspects"¹⁵ (Clark & Gerrig 1990), passe par une

¹⁴ C'est-à-dire "l'ensemble des représentations publiquement partagées qui s'élabore [...] coopérativement au long d'un discours" (Groupe de Fribourg 2012: 22), correspondant plus ou moins ailleurs à la notion de *common ground*.

¹⁵ Pour Clark et Gerrig (1990), cela signifie que le sujet communicant a l'intention que ces aspects soient reconnus comme décriptifs, qu'ils représentent un objet par ressemblance.

modulation du signal, que ce soit au niveau du rythme, de la hauteur intonative, de l'intensité ou de la qualité de voix. Toutes ces caractéristiques participent ainsi à l'attribution de R dès lors qu'elles construisent une énonciation comme distinctive, décalée de l'énonciation en cours et attribuable à une source, en indexant par exemple certaines propriétés relativement stables de son identité, comme sa provenance géographique, son âge, sa classe sociale ou son sexe. Mais le corps du sujet parlant est également un corps affectif et affecté, un corps sensible aux modifications incessantes de son environnement (quel qu'il soit), et qui s'y adapte en permanence. Or c'est également à l'indication de ces transformations que peut contribuer la mise en voix lorsqu'elle signale "the speech activities and the affective stance of the reported characters" (Günthner 1999: 704). Cela relève toutefois davantage du *message*, sur lequel nous allons concentrer désormais notre attention, en conjonction avec l'*attitude*

4.3 Le message et l'attitude

Comme nous l'avons dit, la vocalisation contribue à la construction même du message, dans le sens d'une caractérisation de l'activité communicative et du positionnement affectif (cf. Günthner 1999, ci-dessus) du locuteur ou de la locutrice représenté·e. Voyons maintenant un autre exemple, tiré d'une conversation entre trois ami·e·s, et portant sur la visite de L2 d'un prestigieux magasin de meubles en vue d'acheter un lit:

(2)

- 1 L1: mais euh elle enlève ses lunettes d'un air dramatique tu sais ça
faisait vraiment film euh limite tremblant pis **elle fait** (0.4) °↓bon (0.4)
mille deux cents francs vous comptez!°
2 L2: are you fucking kidding me
3 L3: [((rire))]
4 L1: [à peu] près mais dans d'autres termes (1.1) **elle a fait** (0.2)
°↓mais j/ je pourrai pas vous satisfaire!°
5 L3: ((rire))

(unifr14a05d)

Dans le premier tour de parole (TP), L1 verbalise déjà quelques éléments jugés pertinents d'une scène représentée, en l'occurrence de gestualité exprimant une attitude. Il procède ainsi d'abord à une *description* de la locutrice représentée, dans la typologie clarkienne. Est intéressante ensuite la comparaison avec le cinéma: le locuteur semble bien conscient du "petit drame" qu'il est en train d'installer, en convoquant un univers de référence qui se définit fondamentalement par le "make-believe". Puis L1 poursuit sur sa lancée avec l'énonciation et la vocalisation de R, introduite par le quotatif "faire", dont le signifié actionnel très large incline à l'associer à une représentation multimodale¹⁶, en contraste avec le signifié d'autres verbes de parole qui sélectionnent un objet de nature prioritairement verbale, comme "dire". Au moment d'actualiser l'énonciation représentée, L1 diminue le rythme

¹⁶ Voir Fiedler (2022).

d'allocution et la hauteur intonative et adopte une voix quasi chuchotée, mettant en scène une instance un peu hésitante et contrite. Cette vocalisation contribue d'un côté à l'*attribution*, à l'élaboration d'une *persona* particulière de vendeuse très policée, révérencieuse, qu'on tend à trouver dans les enseignes de luxe; et de l'autre, à la dépeinture de son attitude confuse, due à la gestion délicate de paramètres contextuels contradictoires: rester polie et préserver la face de son interlocuteur, tout en effectuant l'acte illocutoire menaçant de signaler l'incapacité de satisfaire sa requête, en renvoyant le destinataire à son statut économique inadapté au magasin¹⁷.

Le second TP présente aussi des aspects notables. L2 module son énonciation d'une voix très grave et monocorde, dans un rythme cette fois martelé, qui contraste avec celui hésitant de la locutrice représentée dans le TP précédent. Ce faisant, il met en scène une énonciation alternative à celle représentée par L1, où s'effectue d'une certaine façon la même action communicative que celle actualisée précédemment, mais sans le vernis de politesse, bien au contraire. L2 joue une *persona* sur-masculinisée et implacable, en anglais, dans un registre plutôt familial voire vulgaire (le "fucking"), dont l'énonciation tient lieu de réponse négative à la requête du client, mais au moyen d'une question rhétorique plutôt agressive¹⁸. C'est-à-dire que L2 participe à la mise en scène de L1, mais en attribuant à la vendeuse, par la mise en voix et en mots du R concurrent, une *persona* et une attitude tout à fait différentes, et très invraisemblables vu la situation représentée. Ce parasitage énonciatif produit des effets dramatiques particuliers, d'une part en travestissant la locutrice représentée en espèce de butor peu soucieux de la préservation mutuelle des faces, et d'autre part en présentant cette énonciation comme exprimant hypothétiquement la position authentique de la locutrice représentée, derrière le masque de socialisation. C'est d'ailleurs l'évaluation que fait L1 de cette version alternative, en la décrivant comme une reformulation ("en d'autres termes") de celle qu'il construit lui-même. Par ailleurs, cette représentation de L2 convoque un type assez bien délimité, formé de l'appariement d'une séquence verbale et d'une attitude, dans certaines situations particulières¹⁹. La formule "are you fucking kidding me" est en effet identifiée comme un *mème*, et figure dans la base de données <https://knowyourmeme.com/>, où il est défini comme "a rage comic character wearing a scornful facial expression that is

¹⁷ Sur la gestion interactionnelle des faces et la politesse, voir Goffman ([1959] 1990, [1967] 1982), Brown & Levinson (1987), Kerbrat-Orecchioni (2005), *inter alia*.

¹⁸ C'est-à-dire que si le *contenu* est relativement similaire dans les deux propositions d'énonciation représentée de la vendeuse, la *relation* s'avère elle diamétralement opposée.

¹⁹ L'*Urban Dictionary* commente l'énoncé "Are you fucking kidding me?" de cette façon: "You will say this when you question what is asked or told to you in the most extreme way possible." (<https://www.urbandictionary.com/define.php?term=Are%20you%20fucking%20kidding%20me%3F>)

typically used in reaction to someone else's extreme stupidity"²⁰. Cela illustre bien l'hétérogénéité sémiotique pouvant composer une dépicition: une mise en mots, en particulier ici en termes de langue et de registre; une mise en voix particulière; et une expression faciale, dans le cas du même. Les deux derniers ont un rapport étroit à la corporalité, étant donné qu'ils expriment une réaction physiologique à un stimulus particulier. Et en même temps, la convocation d'un type identifiable et identifié donne une idée de la manière dont les voix circulent et se configurent dans la société, lorsqu'elles en arrivent à s'organiser en constellations préconstruites, actualisant de façon concomitante une *persona* et une attitude dans une situation particulière.

L'exemple suivant nous permettra d'illustrer davantage l'intrication entre *attribution*, *message* et *attitude*, ainsi que l'interaction entre la mise en voix et en mots dans l'élaboration de ces aspects.

(3)

1 L1: t'en as une elle a cinq ans pis elle pleure tout le temps dès que tu élèves la voix (0.5) **tu dis** qu'elle a fait un truc pas bien (0.3) elle pleure (0.9) ((halètement sanglotant)) #c'est pas moi# t'es % ((rire))
 2 L2: ((rire))\$si je sais que c'est toi\$ [((rire))]
 3 L1: [elle elle] avait glissé (0.2) pendant la soirée de gym tu sais ils couraient dans toute la salle pis **on leur a dit** mais faut pas courir (0.6) mh mh (0.2) pis elle court bien sûr elle tombe avec un autre (0.3) pis je vais vers elle pis **je dis** enfin quand elle s'est enfin consolée **je dis** mais faut pas courir ((halètement sanglotant)) inon bon (0.3) pis tu la recalmes elle repleure dès que **tu lui dis** qu'il fallait pas (0.2) #c'est pas moi c'est lui# oui non mais (0.5) ouais (0.2) bon
 4 L2: ((rire)) elle est traumatisée la pauvre [gamine]
 5 L1: [% elle est] elle est (0.2) tu sais mais vraiment la petite chouchou que je pense que jamais jamais on lui parle plus haut que (1.2)
 6 L2: on lui reproche rien [elle fait tout bien]
 7 L1: [on lui reproche jamais rien quoi] (0.2) mais c'est elle qui s'est levée une fois pour aller boire (0.9) % j'ai soif (0.3) ouais [alors euh.]
 8 L2: [mais c'est ces petites] princesses %

(unine15b03d)

Dans cet exemple, L1 produit une mise en scène fortement investie sur le plan multimodal. Elle représente en effet plusieurs actions – en particulier de la locutrice représentée qui est son élève – au moyen de dépicitions dont la dimension verbale est extrêmement réduite (le "mh mh" ou le halètement sanglotant), mais qui s'inscrivent avec les dépicitions davantage verbalisées à la fois dans une narration, à titre d'événements dans une chaîne causale, et dans une interaction communicative, en tant par exemple que confirmation de réception d'une instruction ("mh mh") ou dénégation ("c'est pas moi") au sein de paires adjacentes. Dans cette scène représentée, la vocalisation contribue à l'*attribution*, avec une mise en place doublement asymétrique, du point de vue

²⁰ <https://tenor.com/fr/search/are-you-fucking-kidding-me-gifs>.

des places relativement stables, avec une différence d'âge (adulte/enfant) et de statut (enseignante/élève), marquée par une représentation plus aiguë de la voix de l'élève; et des places éphémères, mouvantes, liées aux actions communicationnelles des participantes, qui sont bien sûr associées à leur statut respectif. Cela contribue ainsi au *message*, et notamment au signalement de "the speech activities and the affective stance of the reported characters" (Günthner 1999: 704). D'ailleurs, on peut remarquer que la contribution de la vocalisation à l'*attribution* et au *message*, en déterminant qui parle et en délimitant dans le flux communicatif les différentes énonciations, favorise un allègement remarquable de l'introduction des DD. Par exemple le segment "**je dis** mais faut pas courir ((halètement sanglotant)) ↓non bon↓" contient un seul M constitué minimalement de la mention de la locutrice représentée et d'un verbe de parole, et trois R adjacents, dont seul le premier est véritablement encadré. Or l'interprétation n'en est pas pour autant sensiblement affectée.

Ensuite, l'exemple montre bien les combinaisons des moyens communicatifs. L1 construit sa mise en scène en décrivant certaines actions communicatives, dont plusieurs se réduisent à cette verbalisation (par exemple "tu la recalmes"). D'autres fonctionnent en combinaison avec des dépicions, que ce soit à titre de complément d'objet direct d'un verbe de parole comme dans "**on leur a dit** mais faut pas courir"²¹; ou d'"adjunct" (Clark 2016: 326) à une partie de l'énoncé, c'est-à-dire qui s'y attache à la façon d'un modifieur non restrictif, participant à l'élaboration d'un contenu sans être un constituant obligatoire. C'est le cas dans "elle pleure (0.9) ((halètement sanglotant)) #c'est pas moi#". L2 participe d'ailleurs à ce tissage sémiotique dense, en interagissant à la fois par des descriptions ("elle est traumatisée la pauvre gamine") et par des dépicions attribuées à l'enseignante représentée ("si je sais que c'est toi").

Finalement, un tel échange montre bien à quel point un DD s'éloigne d'une reproduction *verbatim*. C'est bien une "construction" (Tannen 1989) subjective, recontextualisée et reconfigurée selon les intentions communicationnelles du locuteur représentant et les contraintes de sa propre situation, de sorte que la question du "how is this 'other voice' being done?" est inextricablement liée à celle du "what is the speaker doing with this 'other voice'?" (Couper-Kuhlen 1998: 9). Cet enchevêtrement du représentant et du représenté est d'autant plus difficilement démêlable que les moyens pouvant être versés au compte de l'*attitude* sont virtuellement infinis, relevant autant du choix des quotatifs que de la description des participants, la mise en voix des énonciations, mais aussi

²¹ C'est bien en tant qu'icône, mention, mime ou démonstration qu'un DD tient lieu de complément d'objet. L'intégration syntaxique relève ainsi comme le note le Groupe de Fribourg d'un "collage intersémiotique" (2012: 68).

l'organisation des TP²² ou la distribution quantitative du temps de parole. Comme le souligne Authier-Revuz, "le locuteur qui cite: pas un magnétophone!" (1992: 16). Autrement dit: "L'effacement ostentatoire de L au seuil de l'énoncé qu'il rapporte est un simulacre de non-intervention absolue de L dans son acte de rapport" (Authier-Revuz 1978: 51), qui tient à un effet de sens du mode sémiotique de la monstration, ou vu différemment, à l'inscription de l'énonciation dans une dépicition. Günthner va dans ce sens, en convoquant l'image très bakhtinienne de la réfraction:

The recontextualized utterances are stylized, exaggerated, and caricatured and are made to accommodate the reporter's evaluations. These kinds of polyphonic utterances, simultaneously expressing both the intention of the character who is being quoted, and the refracted intention of the reporter, are not restricted to literary texts, but are frequently used in everyday interactions as well. (Günthner 1999: 696).

Ainsi, là où l'évaluation verbale univoque du locuteur représentant se limite à M, c'est-à-dire à son propre territoire énonciatif, le type d'évaluation camouflée dont nous venons de parler fonctionne sur le mode du "en même temps"²³, le locuteur investissant de toutes parts le territoire de l'autre, tout en présentant comme autonome, indépendante, une énonciation qu'il a faite sienne, résultat d'un jeu de miroirs complexe.

5. Conclusion

Nous espérons avoir suggéré quelques pistes productives dans la prise en considération de la mise en voix (parmi les autres ressources du répertoire sémiotique) affectée au DD, à la fois dans l'examen de sa contribution spécifique à l'actualisation de l'une ou l'autre composante du DD en fonction d'affordances particulières, et dans l'analyse des interactions possibles avec d'autres ressources sémiotiques. Les exemples d'interactions ont montré la densité du feuilletage énonciatif, chaque énonciation pouvant être investie d'une superposition complexe de voix et d'intentions, dont l'attribution et la caractérisation sont loin d'être évidentes, en particulier lorsqu'on envisage le mode de communication analogique. Pourtant, étant donné la façon dont ce dernier semble se prêter à "traduire toutes les nuances possibles de la distance de Loc aux paroles qu'il emprunte" (Authier-Revuz 1978: 70), son intégration plus poussée dans l'étude du DR en général nous semble absolument essentielle.

²² Certains TP semblent avoir une importance stratégique supérieure à d'autres; par exemple il y a souvent un avantage à qui a le "dernier mot", et dans notre exemple, c'est dans la majorité des cas l'enseignante.

²³ On le retrouve tel quel dans la définition du DR de Vološinov (1929) déjà mentionnée, ainsi que dans la citation de Günthner juste en dessus, avec le "simultaneously".

REFERENCES

- Atlani, F. (1984). On l'illusionniste. In A. Grésillon & J. L. Lebrave (éds.), *La langue au ras du texte* (pp. 13-29). Lille: Presses Universitaires de Lille.
- Authier-Revuz, J. (1978). Les formes du discours rapporté. Remarques syntaxiques et sémantiques à partir des traitements proposés. *Documentation et Recherche en Linguistique Allemande Vincennes*, 17, 1-87.
- Authier-Revuz, J. (1992). Repères dans le champ du discours rapporté. *L'Information Grammaticale*, 55, 38-42.
- Authier-Revuz, J. (2020). La représentation du discours autre. *Principes pour une description*. Berlin/Boston: de Gruyter
- Avanzi, M., Béguelin, M.-J., Corminboeuf, G., Diémoz, F. & Johnsen, L.A. (2023). *OFROM – corpus oral de français de Suisse romande*. Université de Neuchâtel, <https://ofrom.unine.ch/>.
- Bakhtine, M. (1978). *Esthétique et théorie du roman*. Paris: Gallimard.
- Bavelas, J. B., Gerwing, J., & Healing, S. (2014). Effect of dialogue on demonstrations: Direct quotations, facial portrayals, hand gestures, and figurative references. *Discourse Process*, 51, 619-655. <https://www.tandfonline.com/doi/abs/10.1080/0163853X.2014.883730>
- Becker, A. L. (1991). Language and languaging. *Language & Communication*, 11(1), 33-35. [https://doi.org/10.1016/0271-5309\(91\)90013-L](https://doi.org/10.1016/0271-5309(91)90013-L)
- Brown, P. & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Caillat, D. (2012). Le discours représenté direct dans les conversations orales: une gestion multimodale des décrochements énonciatifs [en ligne]. *Le Discours et la Langue*, 2(2), 113-122. <https://hal.science/hal-00814750/document>
- Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, 123(3), 324-347. <https://doi.org/10.1037/rev0000026>
- Clark, H. H. & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, 66(4), 764-805. <https://doi.org/10.2307/414729>
- Couper-Kuhlen, E. (1998). Coherent voicing. On prosody in conversational reported speech. *Interaction and Linguistic Structures*, 1, 1-28.
- de Cornulier, B. (1978). L'incise, la classe des verbes parenthétiques et le signe mimique. *Cahier de linguistique*, 8, 53-95. <https://doi.org/10.7202/800060ar>
- Détie, C. (2008). Textualisation et (re)conditionnement énonciatif. In *Congrès Mondial de Linguistique Française 200*, 1331-1344. Les Ulis: EDP Sciences. <https://doi.org/10.1051/cmlf08127>
- Ducrot, O. (1984). *Le dire et le dit*. Paris: Minuit.
- Fiedler, S. (2022). Une reconsidération du discours rapporté en langue parlée avec être là, faire (à quelqu'un) et se dire. *Langue française*, 216, 29-46.
- Fløttum, K., Jonasson, K. & Norén, C. (2007). *ON: pronom à facettes*. Bruxelles: De Boeck.
- Fox, B. A. & Robles J. (2010). It's like *mmm*: Enactments with *it's like*. *Discourse Studies*, 12, 715-738.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Goffman, E. ([1959] 1990). *The presentation of the self in everyday life*. Londres: Penguin.
- Goffman, E. ([1967] 1982). *Interaction ritual. Essays on face-to-face behavior*. New York: Pantheon.
- Goffman, E. (1981). *Forms of talk*. Philadelphie: University of Pennsylvania Press.
- Groupe de Fribourg. (2012). *Grammaire de la période*. Berne: Peter Lang.

- Günthner, S. (1999). Polyphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics*, 31(5), 685-708. [https://doi.org/10.1016/S0378-2166\(98\)00093-9](https://doi.org/10.1016/S0378-2166(98)00093-9)
- Hanote, S. & Chuquet, H. (2004). *Who's speaking, please? Le discours rapporté*. Paris: Ophrys.
- Hodge, G. & Cormier, K. (2019). Reported speech as enactment. *Linguistic Typology*, 23, 185-196. <https://www.degruyter.com/document/doi/10.1515/lingty-2019-0008/html>
- Hodge, G. & Johnston, T. (2014). Points, depictions, gestures and enactment: Partly lexical and non-lexical signs as core elements of single clause-like units in Auslan (Australian Sign Language). *Australian Journal of Linguistics*, 34(2), 262-291. <https://www.tandfonline.com/doi/abs/10.1080/07268602.2014.887408>
- Hsu, H.-C., Brône, G. & Feyaerts, K. (2021). When gesture "takes over": Speech-embedded nonverbal depictions in multimodal interaction. *Frontiers in Psychology*, 11, 1-23. <https://doi.org/10.3389/fpsyg.2020.552533>
- Jacquin, J. (2017). Le pronom ON dans l'interaction en face à face: une ressource de (dé)contextualisation. *Langue française*, 193(1), 77-92. <https://doi.org/10.3917/lf.193.0077>
- Kerbrat-Orecchioni, C. (2005). *Le discours en interaction*. Paris: Armand Colin.
- Koch, P. & Oesterreicher, W. (2001). Langage parlé et langage écrit. In G. Holtus, M. Metzeltin & C. Schmitt (éds.), *Lexikon der Romanistischen Linguistik (Band I, 2)* (pp. 584-627). Berlin/Boston: de Gruyter.
- Kusters, A., Spotti, M., Swanwick, R. & Tapio, E. (2017). Beyond languages, beyond modalities: Transforming the study of semiotic repertoires. *International Journal of Multilingualism*, 14(3), 219-232. <https://doi.org/10.1080/14790718.2017.1321651>
- Mahrer, R. (2014). Écrire et parler. Quelques préalables théoriques. *Genesis. Manuscripts – Recherche – Invention*, 39. <https://doi.org/10.4000/genesis.1371>
- Mahrer, R. (2019). Parler, écrire: "continuum communicatif" et rupture matérielle. *Pratiques*, 183-184. <http://journals.openedition.org/pratiques/6842>.
- Maingueneau, D. (2002). Problèmes d'ethos. *Pratiques*, 113(1), 55-67. <https://doi.org/10.3406/prati.2002.1945>
- Maingueneau, D. (2012). Un réseau de concepts. Entretien avec Dominique Maingueneau au sujet de l'analyse du discours littéraire, propos recueillis par Reindert Dhondt et David Martens. *Interférences littéraires/Littéraire interferences*, 8, 203-221.
- Maingueneau, D. (2020). Hiérarchie d'ethos et locuteurs typifiés. E-rea. *Revue électronique d'études sur le monde anglophone*, 17(2), en ligne. <https://doi.org/10.4000/erea.10001>
- Park, Y. (2009). Interaction between grammar and multimodal resources: Quoting different characters in Korean multiparty conversation. *Discourse Studies*, 11(1), 79-104.
- Peirce, C. S. (1978). *Écrits sur le signe*. Paris: Seuil.
- Rosier, L. (2008). *Le discours rapporté en français*. Paris: Ophrys.
- Sidnell, J. (2006). Coordinating gesture, talk, and gaze in reenactments. *Research on Language and Social Interaction*, 39(4), 377-409. https://doi.org/10.1207/s15327973rlsi3904_2
- Spronck, S. (2022). Voloshinov's thesis: Towards a positive collaborative research programme on reported speech. Paper presented at the Workshop *Cross-disciplinary perspectives on quoting and speech reporting*, Université libre de Bruxelles.
- Spronck, S. & Nikitina, T. (2019). Reported speech forms a dedicated syntactic domain. *Linguistic Typology*, 23(1), 119-159. <https://doi.org/10.1515/lingty-2019-0005>
- Stec, K., Huiskes, M. & Redeker, G. (2015). Multimodal analysis of quotation in oral narratives. *Open Linguistics*, 1(1). <https://doi.org/10.1515/opli-2015-0018>

- Tannen, D. (1989). *Talking voices: Repetition, dialogue and imagery in conversational discourse*. Cambridge: Cambridge University Press.
- Thompson, G. (1996). Voices in the text: discourse perspectives on language reports. *Applied Linguistics*, 17(4), 501-530. <https://doi.org/10.1093/applin/17.4.501>
- Vion, R. (1995). La gestion pluridimensionnelle du dialogue. *Cahiers de Linguistique Française*, 17, 179-203.
- Vološinov, V. N. ([1929] 2010). *Marxisme et philosophie du langage: Les problèmes fondamentaux de la méthode sociologique dans la science du langage* (traduction I. Tylkowski-Ageeva & P. Sériot; nouvelle édition bilingue). Limoges: Lambert-Lucas.
- Watzlawick, P., Beavin, J. H. & Jackson, D. D. (1972). *Une logique de la communication*. Paris: Seuil.
- Winkin, Y. ([1981] 2000). *La nouvelle communication*. Paris: Seuil.

Annexe

Conventions de transcription:

- (.): pauses d'une durée inférieure à 0.2 secondes
- (0.n): pauses chronométrées en secondes
- [...]: chevauchements
- (...): vocalisations non-lexicales
- %: productions incompréhensibles
- ↑ ↓: montées et descentes intonatives
- \$...\$: segments vocalisés en riant
- #...#: segments avec une intonation peinée