Revue COSSI

**N° 15 | 2025**
**Entre sciences de l’information et de la communication et sciences cognitives : les bases d’un dialogue possible**

# Entre production audiovisuelle multimodale et réception contrastée d’une chronique politique sur YouTube, quels enjeux ?

***Louise OUDJANI*** *Docteur*
*Sciences de l'information et du document*
*Geriico*
*University of Lille*

**Édition électronique :**
**URL :**
https://revue-cossi.numerev.com/articles/revue-15/3498-finalentre-production-audiovisuelle-multimodale-et-reception-contrastee-dune-chronique-politique-sur-youtube-quels-enjeux
**DOI :** numerev_2759
**Date de publication :** 21/06/2025

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : OUDJANI, L. (2025) Entre production audiovisuelle multimodale et réception contrastée d’une chronique politique sur YouTube, quels enjeux ?. *Revue COSSI*, (15). https://doi.org/10.34745/numerev_2759

Cet article examine la multimodalité dans les vidéos politiques sur YouTube à partir de l’analyse de la chronique Ouvrez les guillemets d’Usul, diffusée sur la chaîne de Mediapart entre 2017 et 2021. La méthodologie combine une analyse sémiotique et communicationnelle de 78 épisodes et neuf entretiens compréhensifs avec des spectateurs. Les résultats mettent en évidence trois référentiels principaux qui s’articulent pour produire un discours politique critique et intertextuel. La réception apparaît contrastée : elle suscite consensus et engagement chez les spectateurs familiers des codes, mais reste conditionnée par leur bagage culturel. La discussion souligne les ambivalences de cette multimodalité et l’importance de développer la littératie visuelle en éducation aux médias afin de réduire les inégalités face aux stratégies de persuasion multimodale en ligne.

---

This article examines multimodality in political videos on YouTube based on the analysis of Ouvrez les guillemets, broadcast on the Mediapart YouTube channel between 2017 and 2021. The methodology combines a semiotic and communicational analysis of 78 episodes with nine in-depth interviews with viewers. The findings highlight three main frameworks that interact to produce a critical and intertextual political discourse. Audience reception appears mixed: it generates consensus and engagement among viewers familiar with the codes, but remains shaped by their cultural background. The discussion emphasises the ambivalences of this multimodality and stresses the importance of developing visual literacy within media education in order to reduce inequalities in the face of online multimodal persuasion strategies.

---

**Mots-clés :**
Réception, Politique, Publics, Multimodalité, YouTube, Vidéo, Polysémie de l'image, Littératie visuelle

---

# Introduction

Sur YouTube, l’image politique constitue un objet d'étude protéiforme marqué par l’entrelacement de registres médiatiques, culturels et musicaux qui participent à la construction du discours. Cette hybridité se retrouve dans la chronique Ouvrez les guillemets du vidéaste Usul, diffusée sur la chaîne YouTube de Mediapart. À travers les épisodes de la chronique, le vidéaste combine des images d’archives, des séquences extraites de médias classiques et d'autres références, qu'elles soient musicales, numériques ou cinématographiques. L’image politique sur YouTube apparaît ainsi comme une forme multimodale, mobilisant simultanément image, son, texte et références intertextuelles.

Dans cet article, la notion de multimodalité est centrale pour appréhender les enjeux de cette réception. Selon la théorie cognitive de l’apprentissage multimédia (Mayer, 2009 ; Amadieu et Tricot, 2014), la combinaison de canaux visuels et auditifs favorise la mémorisation et la compréhension, tout en posant la question de la charge cognitive (Sweller, 1988). L’insertion de références musicales, culturelles et médiatiques fonctionne comme un système de référentialité, c’est-à-dire comme un ensemble de signaux sémiotiques qui orientent l’interprétation du spectateur en activant des connaissances préalables et des imaginaires collectifs (Aumont, 2020 ; Douyère, 2019). Ces procédés, relativement fréquents dans la culture du remix et du mème (Allard, 2015 ; Bédard et al., 2022), constituent autant de leviers rhétoriques de la production amateure en ligne.

Sur le plan cognitif, la réception de ces contenus renvoie à des mécanismes d’attention sélective, de mémoire associative et d’évaluation de la crédibilité. Les recherches récentes sur l’apprentissage à partir de vidéos montrent que l’engagement du spectateur est modulé par son bagage culturel, ses littératies et sa capacité à évaluer la source. En effet, Delgado et al. (2022) démontrent que les vidéos captent plus facilement l’attention mais ne garantissent pas une meilleure compréhension de l'information. En revanche, les textes permettent une lecture plus profonde et une meilleure structuration de l’information. Autrement dit, la vidéo favorise l’engagement émotionnel, mais peut limiter la profondeur critique. Aussi, Lescarret et al. (2023) montrent quant à eux que les spectateurs filtrent l’information selon leurs croyances et attitudes préexistantes. En étudiant les pratiques estudiantines, les auteurs révèlent que les étudiants montrent une résistance un peu plus grande, mais restent tout de même influencés. Plus tard, en se rapprochant d'un public de collégiens, ces mêmes auteurs mettent en lumière que le simple fait de rappeler aux élèves de vérifier la source améliore leur capacité à juger de la crédibilité d'une information. Dans le contexte des réseaux sociaux numériques, la question de la réception ne se limite pas à la compréhension des messages, mais inclut aussi l’attention fluctuante au sein d'environnements saturés. De Koning et al. (2009) ont montré que la mise en place d'un cadre théorique permet un « guidage attentionnel » au sein des animations pédagogiques. Les résultats révèlent ainsi que les signaux visuels (surlignages, zooms, couleurs) permettent de diriger l’attention et d’améliorer la compréhension du sujet. En somme, en l'absence de guidage, la surcharge cognitive est plus probable. Plus tard, l’étude de Roozenbeek et al. (2022), fondée sur l’analyse de plusieurs plateformes,

examine l’efficacité de l’« inoculation psychologique » pour renforcer la résilience face à la désinformation sur les réseaux sociaux. Les résultats montrent en définitif que les interventions préventives (telles que les jeux, les vidéos éducatives et l’éducation aux médias et à l’information) améliorent la capacité des usagers à reconnaître les techniques de manipulation à l’œuvre. Ces éléments invitent à formuler une problématique au croisement des SIC et des sciences cognitives : comment la multimodalité, dans la chronique Ouvrez les guillemets, contribue-t-elle à la construction du discours politique et comment ces stratégies visuelles et sonores influencent-elles la réception ?

Les enjeux de cette problématique nous amènent à circonscrire trois principales questions de recherche :

- Q1 : quels référentiels médiatiques, culturels et musicaux sont mobilisés dans ces vidéos et comment structurent-ils le discours visuel ?
- Q2 : quelles stratégies de légitimation et de persuasion sont activées par l’usage de références multimodales ?
- Q3 : comment la réception varie-t-elle selon le bagage cognitif et culturel des spectateurs ?

Afin d’apporter des éléments de réponse à ces interrogations, nous inscrivons notre démarche dans un double cadre théorique qui conjugue les apports des sciences de l’information et de la communication, mobilisant les notions d’intertextualité, de rhétorique visuelle et de littératie visuelle, avec ceux des sciences cognitives, qui permettent d’éclairer les phénomènes de charge cognitive, de mémoire associative et d’attention sélective. Cette articulation a pour but de rendre compte à la fois des stratégies sémiotiques déployées dans les vidéos et des mécanismes psychologiques sollicités chez les spectateurs. Dans un second temps, nous détaillerons les choix méthodologiques et présenterons les résultats qui en sont issus. Enfin, ces résultats seront discutés à la lumière de l’état de l'art existant, afin de mettre en évidence les apports de cette recherche, mais également ses limites et les perspectives qu’elle ouvre.

# Un cadre conceptuel à l'intersection des sciences de l’information et de la communication et des sciences

# cognitives

L’image politique sur YouTube s’inscrit dans une culture du remix et de la circulation intermédiatique. Julien Longhi (2018) parle de « genre YouTube » pour désigner des productions régies par des codes visuels et narratifs partagés (humour, connivence, régularité), évalués émotionnellement (like). David Douyère (2019) souligne le rôle central de la remédiatisation : images reprises, transformées et réinsérées. Limor Shifman (2013) montre que le mème est un format culturel fondé sur la répétition et la reconnaissance, tandis que Bédard, Millette et Brouard (2022) en analysent la portée politique, marquée par l’ironie et la satire. Ces pratiques peuvent toutefois exclure les publics non initiés. Elles rejoignent les apports de Roland Barthes (1982) sur l’ancrage de l’image, qui guide l’interprétation. Des travaux récents confirment ces enjeux : Salojärvi (2023) met en évidence la performativité multimodale dans les campagnes politiques, et Törnberg (2025) montre le rôle persuasif de l’esthétique visuelle dans la désinformation climatique.

De plus, l’ensemble de ces pratiques s’inscrit dans la perspective de la littératie visuelle. Définie initialement par Debes (1969) et approfondie par Avgerinou et Ericson (1997), elle désigne la capacité à analyser et interpréter les messages visuels. Dans leurs travaux, Lebrun (2015) et Farmer (2019) insistent d'ailleurs sur ses dimensions esthétiques et socio-culturelles. Plus récemment, des recherches empiriques soulignent les limites persistantes dans l’acquisition de cette compétence. Uminski (2025) a montré, dans une étude expérimentale auprès d’étudiants américains, que la lecture d’images scientifiques reste problématique pour une partie des apprenants. D'ailleurs, dans le contexte de l’enseignement supérieur au Koweït, Zayeb, Aleidan et Ali (2024) notent que les enseignants eux-mêmes manquent de formation à l’analyse visuelle, ce qui limite leur capacité à transmettre ces compétences à leurs étudiants. Ces résultats nous indiquent que la littératie visuelle est inégalement distribuée et conditionne l’aptitude des spectateurs à décoder les référentiels culturels et à construire une réflexion critique de l'information. A cet égard, l’image politique sur YouTube relève d’un jeu d’intertextualité, de remédiatisation et de capitaux (dans le sens bourdieusien du terme).

Les sciences cognitives envisagent les mécanismes de réception des vidéos multimodales sous le prisme de la charge cognitive. Rappelons que John Sweller (1988) a formulé la théorie de la charge cognitive, selon laquelle l’efficacité d’un apprentissage dépend de la gestion entre trois types de charge : intrinsèque (qui est liée à la complexité du contenu), extrinsèque (celle-ci est liée à la présentation de l’information) et essentielle (liée au traitement en mémoire de travail). Appliquée à notre objet d'étude, cette perspective met en exergue le paradoxe de cette multimodalité : la densité des références et la rapidité du montage peuvent stimuler l’engagement mais aussi saturer les ressources cognitives des spectateurs. En 2009, Richard Mayer, dans

sa théorie cognitive de l'apprentissage multimédia, a démontré que la combinaison du canal visuel et du canal auditif favorise la compréhension et la mémorisation, à condition que ces modalités soient agencées de manière cohérente. Puis, Amadieu et Tricot prolongent cette réflexion en 2014 en soulignant que les vidéos ne sont jamais des supports d'apprentissage « transparents » : elles exigent une activité cognitive intense, dont l'efficacité dépend de l'articulation des informations visuelles et sonores. Dans ce cadre, les travaux de De Koning et al. (2009) sur l'importance du guidage attentionnel ajoutent que des signaux explicites (indices visuels, zooms, ralentis, incrustations textuelles) orientent le regard et améliorent la compréhension ; des signaux présents en grand nombre dans chaque épisode de la chronique Ouvrez les guillemets. Ces réflexions scientifiques trouvent des prolongements dans les recherches récentes sur la réception de la multimodalité. En 2025, Liang et al., en étudiant la réception de l'humour multimodal via les « danmu  », montrent que la superposition de signaux textuels, visuels et sonores produit des effets différenciés : un consensus et un amusement partagé pour les spectateurs familiers de ces codes, mais une véritable surcharge cognitive pour les autres. Ce constat rejoint l'hypothèse que la très, ou trop, grande densité de références favoriserait certains publics tout en en excluant d'autres. Le processus mnésique  joue également un rôle important dans cette partition : les vidéos politiques activent des mécanismes de mémoire associative et d'effet de priming . Une chanson issue de la tradition ouvrière, une référence cinématographique ou le logo d'un média évoquent immédiatement un contexte, qu'il soit idéologique, politique ou social, et renforcent par là même l'ancrage du message. Les chercheurs Duan, Peng et Tao confirment en 2025 ce mécanisme en proposant le modèle du « Symbol-Cognition Alignment », selon lequel l'efficacité d'un message multimodal dépend de l'alignement entre symboles visuels et schémas cognitifs préexistants chez le spectateur.

D'autres facteurs de réception différenciée sont mis en avant. En effet, Lescarret, Hoareau et Courbet (2023), puis Lescarret, Courbet et Bernard (2024), ont montré que la crédibilité perçue des vidéos dépend autant des signaux de provenance (logos ou sources institutionnelles) que des dispositions cognitives et affectives des spectateurs. Li (2024), dans une enquête auprès de plus de 2 000 internautes chinois, démontre en outre que la « connexion multimodale » aux médias politiques stimule la participation citoyenne via une chaîne de médiations : une attention portée aux informations, des discussions interpersonnelles, une confiance institutionnelle. Ces résultats soulignent que la multimodalité n'agit pas seulement comme un amplificateur cognitif, mais aussi comme un levier de mobilisation sociale. Enfin, la dimension émotionnelle reste primordiale. Comme le rappelle Joffe (2007), l'image est un vecteur affectif puissant, capable de déclencher des réactions immédiates parfois plus fortes que celles suscitées par le texte. Or cette intensité émotionnelle rend les spectateurs plus vulnérables aux stratégies de persuasion, comme l'indiquent les travaux de Roozenbeek, Van der Linden et Nygren en 2022. L'état de l'art existant montre bien que la réception des vidéos ne se limite pas à la simple compréhension rationnelle de son contenu : en réalité, elle engage un ensemble complexe de processus attentionnels, mnésiques, émotionnels et évaluatifs. Le croisement entre sciences de l'information et de la communication et

sciences cognitives permet d'appréhender la multimodalité politique en tenant compte de l'ensemble de ces facteurs.

# Une approche qualitative de la méthodologie

Cette recherche repose sur une approche qualitative qui articule deux volets complémentaires : l'analyse sémiotique et communicationnelle d'un corpus audiovisuel (n = 78) et la conduite d'entretiens compréhensifs menés auprès des spectateurs de ces vidéos. Cette approche double permet de mettre en relation les stratégies visuelles et narratives déployées par le vidéaste en les mettant en tension des processus de réception, de compréhension et d'évaluation. La thèse, dont proviennent ces résultats, a permis d'étudier de nombreux corpus et nous souhaitons mettre en lumière une partie de ces résultats dans cet article, en mettant l'accent sur la notion de multimodalité.

## Analyse audiovisuelle

Le premier corpus est constitué de 78 vidéos extraites de la chronique Ouvrez les guillemets, diffusée entre 2017 et 2021 sur la chaîne YouTube de Mediapart. Le choix de ce corpus tient à plusieurs raisons : d'une part, il s'agissait d'une production régulière et identifiable dans l'écosystème politique de la plateforme, ce qui en faisait un terrain empirique cohérent pour analyser l'évolution d'une écriture audiovisuelle. D'autre part, le caractère hybride de la proposition était pertinent dans le champ de l'analyse des productions socionumériques : la production amateure du vidéaste Usul se conjuguait de plus à l'autorité journalistique d'un média reconnu (Mediapart). Enfin, c'est son caractère multimodal qui en faisait un objet d'étude à privilégier pour examiner la manière dont la multimodalité structure un discours politique.

L'analyse s'appuie sur une grille inspirée de la sémiotique de l'image (Barthes, 1982 ; Aumont, 2020), de l'approche sémiopragmatique (Gervereau, 1990) et des travaux en SIC sur les écritures numériques (Douyère, 2019 ; Longhi, 2018). Pour Roland Barthes, l'image fonctionne à travers un système de signifiants dont l'ancrage (par exemple par le texte ou les logos) oriente la lecture. Jacques Aumont souligne quant à lui le rôle de la construction visuelle et de l'iconographie dans l'expérience spectatorielle. Ces apports permettent d'examiner comment le montage, les choix typographiques ou les palettes chromatiques guident l'interprétation. De la même façon, ces clés d'analyse de l'image nous permettent d'envisager la logique d'intertextualité des vidéos d'Usul, qui

s’appuient sur des références télévisuelles, cinématographiques ou musicales pour créer une certaine analyse de la politique et de l'actualité. Enfin, les travaux sur la culture du mème (Shifman, 2013 ; Bédard et al., 2022) éclairent l’usage humoristique et satirique des images remixées. Ces procédés favorisent la viralité de l'information mais posent aussi la question d’un entre-soi : la compréhension des références suppose un capital culturel partagé. L’analyse du corpus met donc en évidence les dimensions esthétiques, rhétoriques et stratégiques de l’image politique sur YouTube telles qu'elles sont développées à dessein par le vidéaste Usul.

# Entretiens compréhensifs

Le deuxième corpus repose sur neuf entretiens semi-dirigés compréhensifs menés auprès de spectateurs de la chronique. Le choix de cette méthode se fonde sur les apports de Jean-Claude Kaufmann (1996), qui conçoit l’entretien compréhensif comme une démarche permettant d’accéder à la subjectivité des enquêtés en leur donnant la possibilité d’exprimer librement leur expérience ou leur point de vue. Les participants ont été sélectionnés sur le réseau social Facebook, plus enclin à fournir des coordonnées de prises de contact. L’échantillon reflète une certaine catégorie socioprofessionnelle élevée et moyenne, et un certain nombre d'entre eux travaillent dans la communication ou la culture :

**Figure 1.**

*Présentation des personnes interrogées (n = 9)*

| Anonymés | Sexe | Âge | Métier | Localisation | Dernier diplôme |
|---|---|---|---|---|---|
| Marie-Diane | F | 30 - 35 | Journaliste | Grand-Est | Master Info-Com |
| Léonie | F | 30 - 35 | Journaliste | Rhône-Alpes | Master Journalisme |
| Nicolas | M | 35 - 40 | Bibliothécaire | Bourgogne | Licence Lettres |
| Yann | M | 35 - 40 | Musicien et scénariste | Albanie | Licence Info-Com |
| Maxime | M | 25 - 30 | Professeur certifié | Grand-Est | Master MEEF Histoire |
| Pierre | M | 30 - 35 | Professeur certifié | Ile-de-France | Master SES |
| Malik | M | 30 - 35 | Ingénieur | Grand-Est | Master d'Ingénieur |
| Charlotte | F | 25 - 30 | Professeure certifiée | Grand-Est | Master MEEF Anglais |
| Arthur | M | 40 - 45 | Assistant d'artistes | Ile-de-France | DNSEP |

La grille d'entretien a été conçue de manière à interroger à la fois les pratiques des personnes rencontrées, leur usage de YouTube et leur trajectoire personnelle vis-à-vis de la politique. Parallèlement, une partie spécifique de la grille a porté sur leur réception de la chronique étudiée, en les invitant à s'exprimer sur des aspects tels que la qualité de la vidéo, sa construction ou son intelligibilité. C'est sur la base de ces derniers éléments, directement liés à la réception de la vidéo, que nous présentons les résultats analysés dans cet article.

# Résultats de cette analyse croisée

## Référentiel multimodal

L'examen des épisodes de la chronique Ouvrez les guillemets montre que la

multimodalité est au cœur de la mise en scène du discours politique. Trois grands référentiels structurent ce dispositif : le musical, le citationnel et le médiatique. Ces derniers s'entrecroisent pour produire une certaine esthétique, assez révélatrice des écritures socionumériques en ligne.

## Le référentiel musical

- Des pistes musicales adaptatives : Usul recourt régulièrement à des musiques libres de droit, notamment celles du compositeur américain Kevin MacLeod (assez reconnaissables pour les spectateurs familiers de la culture Internet ; Malik connaît très bien ce compositeur par exemple). Ces pistes sonores soutiennent en arrière-plan les séquences explicatives, et la tonalité musicale (majeure ou mineure) s'adapte au ton ironique ou didactique des éléments du discours.
- Des effets sonores ponctuels : certains bruitages marquent des ruptures ou renforcent l'effet comique. Ainsi, dans un épisode sur la fermeture des services publics, un grincement de porte accompagne la disparition d'un bureau de poste. Des jingles récurrents pastichent les chaînes d'information et en parodient les codes.
- Références culturelles : le corpus révèle une grande variété de musiques référentielles :
- Les chants militants, tels que le Chant des partisans ou des chants russes, ponctuent certains épisodes consacrés aux luttes sociales ou sont associés à des images de manifestations.
- La chanson française (Renaud, Brassens ou encore Balavoine) est mobilisée pour évoquer un imaginaire populaire et contestataire de la culture française.
- Les musiques de jeux vidéo (Final Fantasy, Super Mario) ou d'anime japonais (Great Teacher Onizuka) agrémentent aussi la chronique. Dans un épisode consacré à l'Éducation nationale et à la précarité des enseignants, le thème de Great Teacher Onizuka est utilisé par exemple. L'œuvre est une critique du système éducatif japonais, considéré comme rigide et compétitif. Le titre ou la provenance de la musique ne sont pas mentionnés dans les crédits de l'épisode, à charge pour le spectateur de reconnaître le thème du manga et de comprendre l'intérêt de son application.

## Le référentiel citationnel

Ce référentiel-ci remplit une double fonction : instruire et légitimer, en s'appuyant sur des autorités intellectuelles et militantes, suggérées par le vidéaste comme des références volontairement invisibilisées par les médias dits traditionnels.

- La parole citée : les épisodes intègrent fréquemment des références théoriques et intellectuelles. Usul et Cotentin citent le sociologue Pierre Bourdieu (La Misère du monde), l'économiste Joseph Stiglitz (critique de la mondialisation), l'historienne Michelle Zancarini-Fournel (mouvements sociaux), l'économiste Frédéric Lordon (économie critique) ou encore le militant et avocat Juan Branco (critiques institutionnelles). Ces références apparaissent souvent visuellement. Cela apporte au discours du vidéaste une forme de légitimation savante : son analyse se retrouve adossée à des figures scientifiques et académiques reconnues.
- La parole donnée : au-delà des citations d'ouvrages, certains épisodes intègrent directement dans le corps de la vidéo des interventions d'acteurs extérieurs (journalistes, chercheurs ou militants), enregistrés et réinsérés au montage. Par exemple, des extraits d'interviews de militants associatifs ou de chercheurs en sciences humaines et sociales reconnus dans leur domaine apportent une polyphonie à l'analyse politique d'Usul.

## Le référentiel médiatique

- La remédiatisation est omniprésente dans les épisodes de la chronique, en particulier les extraits de médias traditionnels et des éléments issus de la culture Internet (mèmes célèbres, etc.). Ce référentiel met en lumière la posture critique du vidéaste vis-à-vis des médias dominants, en les détournant, et sa mobilisation volontaire de symboles partagés.
- Extraits télévisés : dans un épisode consacré à Emmanuel Macron, un extrait de discours présidentiel est suivi d'un montage où la parole est ralentie et immédiatement mise en contraste avec un mème visuel circulant en ligne sur le Président pour le disqualifier et le tourner en ridicule.
- Culture web : des mèmes, des GIFs et des images détournées sont incrustés dans le montage de la vidéo.
- Symboles culturels français : les vidéos réutilisent des images d'archives de manifestations ouvrières (afin de représenter la tradition française de la lutte syndicale), des musiques à forte connotation symbolique (des extraits de pistes d'accordéon ou de bal musette pour évoquer les coutumes françaises ou le « vivre à la française »). Des choix qui contribuent à ancrer à dessein le discours dans une mémoire collective de gauche.
- Détournement ou parodie : reprise, pastiche ou parodie des éléments visuels, discursifs d'une œuvre littéraire, musicale, télévisuelle ou symbolique.

## Démonstration et explication des images parodiées (quelques

**exemples)**

La remédiatisation des images nous permet d'appréhender la formation des êtres « visuellement lettrés » (Lebrun, 2015). Une terminologie fort à propos car en étant détournées de leur contexte initial et réinsérées dans de nouveaux cadres discursifs, les images acquièrent en tout état de cause une polysémie renouvelée. Dans ce contexte, les réseaux sociaux numériques sont des espaces privilégiés de production d'imageries socialement constituées et collectivement élaborées. David Douyère (2019) poursuit en décrivant de surcroît la fonction discursive de l'image, qu'il nomme « l'image conversationnelle », en reprenant l'expression d'André Gunthert (2014), dans son article L'image dont on parle, l'image avec laquelle on parle (2019). Il précise que :

**« [...] la pensée ou ce qui deviendra l'énoncé peut être suscité à l'interlocuteur d'abord sous forme d'images, parce qu'il s'efforcera d'expliciter par le langage, ou parce que l'image peut être une façon de mémoriser des pensées, informations et discours associés (Carruthers, 2002). Son propre discours peut provoquer, susciter ou nourrir chez son interlocuteur des images mentales (sans parler du fait que l'objet de la conversation peut être lui-même une image). L'image est donc d'emblée présente dans la conversation. Elle devient elle-même une image. »**

L'usage des couleurs joue un rôle central dans la construction de l'émotion. Comme le souligne Jacques Aumont (2020), les tons froids instaurent une atmosphère mélancolique ou anxiogène, tandis que les couleurs chaudes accentuent l'intensité dramatique. Les codes visuels du cinéma influencent également les productions en ligne. Ainsi, des procédés issus du film noir (éclairages contrastés, angles obliques, jeux d'ombres) sont mobilisés dans la chronique d'Usul afin d'installer une tension narrative, notamment dans l'épisode 11 consacré à Gérard Collomb, dont le générique fait référence à l'émission Faites entrer l'accusé.

**Figure 2.**

*Extrait de l'épisode 11, à 00:32.*



**Figure 3.**

*Le logo de l'émission Faites entrer l'accusé*.




Le choix typographique est également porteur de sens, certaines polices étant associées à des univers culturels spécifiques. Par exemple, la police gothique renvoie aux genres de l'horreur ou du fantastique. Ces conventions graphiques structurent la réception en mobilisant des référentiels partagés. L'épisode 36 d'Ouvrez les guillemets introduit ainsi le personnage de Lorraine, journaliste caricaturale issue de l'univers des chaînes d'information en continu. Marquant le début d'une nouvelle saison, cet épisode adopte de nouveaux codes éditoriaux et propose une critique de la pression exercée par ces médias pour produire un contenu sensationnaliste.

**Figure 4.**

*Extrait de l'épisode 36, à 00:38.*




Le prologue de l'épisode 15 adopte une mise en scène inspirée des clips musicaux des années 1980, renvoyant à l'esthétique new wave de groupes tels que Depeche Mode, The Cure ou Indochine. La palette chromatique immerge le spectateur dans un univers visuel évocateur, sans référence explicitée. Usul et Cotentin réinvestissent ainsi les codes de cette période à travers des plans rapprochés et un travail contrasté sur la lumière (rouge sur le visage d'Usul, blanche sur celui de Cotentin) conférant à la

séquence un esthétisme assumé.

**Figure 5.**

*Extrait de l'épisode 15, à 00:03.*




Par ailleurs, la mention « Libre dans sa tête », apparaissant sous le plan rapproché d'Usul, est une référence explicite à la chanson Diego, libre dans sa tête de Michel Berger et interprétée par France Gall en 1981. Plus loin dans la séquence, un extrait du clip Troisième sexe (1985) du groupe Indochine est inséré dans l'épisode. Les spectateurs découvrent alors une piste de danse improvisée au sein du studio de tournage, laquelle est mise en scène sous un filtre violet à l'image, rappelant ainsi la palette chromatique du vidéoclip original d'Indochine.

**Figure 6.**

*Extrait de l'épisode 15, à 09:12.*



**Figure 7.**

*Extrait du clip Troisième Sexe d'Indochine (version de 1985).*



D'autres allusions parsèment la chronique, notamment à travers les codes visuels de différentes identités visuelles, toutes issues d'émissions, de clips musicaux ou de médias populaires (telles que celles de la chaîne Thinkerview dans l'épisode 57 intitulé "A qui profite le deal ?").

Finalement, l'usage récurrent de références visuelles issues des influences politiques, littéraires et culturelles du vidéaste permet de construire une chronique politique fondée sur des référentiels multiples. L'intégration d'images d'archives, contemporaines ou fabriquées, ainsi que de mèmes, enrichit l'analyse tout en l'ancrant dans un cadre idéologique spécifique. Ce bricolage médiatique, situé entre culture vidéoludique et audiovisuelle, participe à la définition d'un discours visuel sur YouTube, mais entretient aussi un entre-soi fondé sur des codes et références parfois peu accessibles aux non-initiés.

# Les entretiens

Les entretiens mettent en évidence une réception contrastée d'Ouvrez les guillemets, la vidéo étant perçue comme plus accessible que l'écrit. Léonie estime que « c'est beaucoup plus facile d'accès », car cela permet de « poser son cerveau », contrairement à la lecture, plus exigeante. Nicolas souligne lui aussi que « lire demande un investissement » de plus en plus difficile à maintenir, et voit dans la vidéo une combinaison efficace d'information pointue, de références et d'un format conversationnel « pas si innocent que ça ». Pierre ajoute que la vidéo facilite l'accès à l'information « avec un peu plus de légèreté » que l'écrit. Ces propos montrent que la vidéo est perçue comme une médiation réduisant la charge cognitive.

La construction audiovisuelle suscite des jugements nuancés. Arthur adopte un regard

critique et dénonce des procédés de montage répétitifs, tels que « des coupes nettes » associées à « un léger zoom », qu'il qualifie de « tics de facilité » rendant de nombreuses vidéos « assez pauvres ». À l'inverse, Charlotte valorise une construction claire et lisible, permettant de percevoir « le squelette de la réflexion », ainsi qu'un dispositif visuel « très épuré ». Elle critique les vidéos saturées d'effets, comme « des phrases qui apparaissent » ou « des flashs de lumière », qui nuisent selon elle à l'attention et à l'appréciation du contenu.
La place de l'image et du son est également soulignée dans la perception de la valeur ajoutée du format. Pour Nicolas, le recours à une grammaire visuelle identifiable constitue un signe de professionnalisme et de sérieux : « L'ajout d'éléments de texte, de panneaux, de petits détails [...] facilite l'identification et l'intérêt. Voir la tête du YouTubeur qui apparaît quelque part, le logo d'une chaîne... ce sont des codes clairs pour la personne qui connaît la chaîne comme pour quelqu'un qui arriverait dessus. » Selon lui, cette signalétique rapproche la vidéo des codes traditionnels du journalisme ou du spectacle, et contribue à crédibiliser la production. Malik met en avant l'importance du visuel pour attirer un public : « Pour attirer son public, ce sera forcément beaucoup plus sur le visuel. [...] Beaucoup en ont littéralement rien à faire du contenu, c'est juste de voir quelque chose qui va les égayer pendant dix minutes. » Cependant, certains spectateurs manifestent une distance critique vis-à-vis des codes visuels dominants sur YouTube. Arthur affirme ainsi que les vignettes au graphisme ostentatoire le rebutent : « En général, les gros titres, les espèces de montages Photoshop avec une figure détournée, je trouve ça d'une laideur... » Il préfère les photogrammes tirés d'œuvres artistiques, qu'il considère plus esthétiques et plus authentiques. Une remarque qui illustre aussi la tension entre une forme d'attractivité visuelle et le rejet d'une standardisation graphique perçue comme artificielle.

Dans l'ensemble, les spectateurs interrogés montrent que la vidéo offre une plus-value par rapport à l'écrit en termes d'accessibilité, de légèreté et de convivialité. Toutefois, cette valeur ajoutée dépend étroitement de la qualité de la construction audiovisuelle. Les attentes portent à la fois sur la clarté du propos, la qualité du son et de l'image, la présence de codes visuels identifiables, mais aussi sur une certaine authenticité esthétique. Les spectateurs de notre panel associent ainsi la fonction de l'image non seulement à la transmission d'un contenu, mais aussi à l'identification d'un style, à la crédibilité d'un discours et à l'expérience de réception.

# Discussion autour d'une intelligibilité sous conditions

L'analyse croisée des 78 épisodes de la chronique Ouvrez les guillemets et des neuf entretiens compréhensifs met au jour une logique d'écriture politique fondée sur la

multimodalité, où s'entrelacent des registres médiatiques, culturels et musicaux. Le caractère composite de la chronique ne se limite pas à des effets de style, il façonne aussi l'intelligibilité du propos et conditionne sa réception. Les référentiels mobilisés par Usul opèrent comme autant de signaux de sens qui densifient le discours. À ce titre, nos résultats vont dans le sens des écritures nativement numériques mobilisées par Julien Longhi (2018). Le chercheur souligne que l'écriture propre à YouTube se caractérise par une forte interactivité et un dialogue implicite entre créateurs de contenu et leurs publics. Les contenus ne sont pas figés : ils circulent, se commentent, se répondent et s'enrichissent à travers les commentaires, les reprises et les remixages, ce qui instaure une forme de récursivité. Parallèlement, cette écriture s'incarne dans des formats hybrides : les vidéos associent des éléments de vlog, de chronique, de reportage ou encore de montage expressif. Ils prolongent également l'analyse de la remédiatisation chez Douyère (2019), en montrant que la reprise, le détournement et la réinsertion d'images dans de nouveaux contextes discursifs ne constituent pas une simple esthétique du recyclage mais une véritable économie sémiotique de la preuve. Chez Usul, les images remédiatisées servent de « preuves par l'image  ». Le fonctionnement de cette économie apparaît nettement dans les verbatim. Nicolas évoque des « incrustations visuelles », des « ralentis », des « détournements » qui « poussent à réfléchir » et laissent entrevoir un montage conçu comme instrument de guidage interprétatif. À cet endroit, nos résultats suggèrent que la chronique d'Usul se situe dans une zone médiane, entre mise à distance critique et mobilisation affective, où l'ironie et la polyphonie de sources coexistent avec des opérations d'ancrage très marquées. On retrouve ici la fonction d'ancrage barthésienne, mais dans un environnement socionumérique où les ancrages sont démultipliés et souvent superposés entre eux.

Cette prolifération de signaux sémiotiques renvoie directement aux mécanismes de réception mis en évidence dans les études cognitives. Plusieurs enquêtés évoquent la densité et la vitesse du montage, Arthur indique devoir « revoir deux fois certains passages », Nicolas observe que « le montage est hyper dense », des formules qui résonnent avec la théorie de la charge cognitive de Sweller (1988) et la théorie cognitive de l'apprentissage multimédia de Mayer (2009). Or, dans Ouvrez les guillemets, les gains attentionnels procurés par le montage et la variété référentielle s'accompagnent d'un coût cognitif visible chez une partie des spectateurs. De Koning et al. (2009) rappellent l'importance de signaux explicites pour guider l'attention. La chronique en abonde, mais l'empilement des repères peut paradoxalement brouiller la hiérarchie des informations lorsque plusieurs niveaux d'indices coexistent à l'écran. Ce point éclaire la divergence des jugements sur la qualité de construction, Charlotte valorise une organisation épurée, « voir le squelette de la réflexion », tandis qu'Arthur critique des « tics de réalisation » devenus standard sur la plateforme, et somme toute perçus comme adjuvants faciles plus que comme véritables dispositifs de clarification.

Sur le plan mnésique et affectif, la convergence entre nos données et l'état de l'art est à signaler. Charlotte insiste sur la prégnance des musiques, « c'est ça qu'on retient le plus longtemps », d'autres comme Arthur décrivent l'importance de lutter avec les

images (les mèmes, en l'occurrence) et par le biais de l'humour, qui sont pour lui des armes. Ces propos s'accordent avec l'idée, chez Joffe (2007), d'une puissance affective spécifique de l'image. Lorsque des référents musicaux et iconographiques s'agrègent à des cadres de sens déjà stabilisés dans la mémoire sociale, l'ancrage est renforcé et la remémoration facilitée. C'est précisément ce que met en œuvre la chronique lorsqu'elle associe divers chants militants, des images d'archives de luttes sociales et reprend des codes graphiques extraits du monde de la télévision ou du web. Mais cette efficacité est à temporiser : Törnberg (2025) montre que l'esthétique visuelle peut aussi devenir un levier de polarisation dans la désinformation, et l'on peut transposer ici l'argument : la qualité plastique et l'ingénierie sonore d'une vidéo politique, même critique, engagent les spectateurs sur un terrain affectif qui peut réduire la distance d'examen. L'enthousiasme de Maxime face à un thème de Final Fantasy qui « parle directement à [sa] génération » est le miroir, pour Yann, d'un décrochage :

**« J'ai des jeunes frères, par exemple. [...] C'est des jeunes qui sont plus dans la vingtaine. Oui, ils partagent beaucoup de choses qu'ils ont trouvé sur YouTube, en lien avec leurs centres d'intérêt, des jeux vidéos et des trucs qui les font marrer, les mangas, ce genre de trucs-là. Puis les gens qui sont plus âgés que moi, dans ma famille, dans mon groupe, partagent beaucoup moins. Voilà, c'est très générationnel. ».**

La médiation affective qui crée de la connivence pour les uns entérine l'exclusion pour les autres, la multimodalité n'unifie pas les réceptions, elle les spécifie. Ces effets différenciés se prolongent lorsqu'on considère la crédibilité perçue. Pierre et Marie-Diane attribuent aux références intellectuelles et aux logos médiatiques un rôle décisif, « ça crédibilise tout de suite », « ça fait sérieux », tandis que Léonie dénonce une accumulation parfois ostentatoire, « presque trop », ce qui rejoint les travaux de Lescarret et al. (2024). Nos matériaux confirment que les indices de sérieux n'agissent pas uniformément, ils interagissent avec l'attente de validation externe, la familiarité avec les codes journalistiques et l'adhésion préalable aux positions défendues. Plus largement, cela rejoint l'enquête de Li (2024) sur la « connexion multimodale » aux médias politiques, où l'attention, la discussion interpersonnelle et la confiance jouent le rôle de médiation. A cet égard, la légitimation par accumulation de références peut certes renforcer l'engagement des personnes interrogées, déjà proches de l'univers culturel du vidéaste, mais susciter aussi une forme de résistance chez d'autres qui y voient une rhétorique de surplomb.

Cet écart nous renvoie aux enjeux de littératie visuelle. Les différences de réception mises en lumière par les entretiens ne relèvent pas seulement d'affinités personnelles, elles manifestent des inégalités de compétences face aux images. Lebrun (2015) insiste sur la littératie visuelle comme capacité à analyser les images dans leurs dimensions esthétiques et socio-culturelles, Farmer (2019) en souligne la portée éducative, Uminski (2025) montre que même des étudiants rencontrent des difficultés d'interprétation visuelle. Or, Ouvrez les guillemets suppose une forte culture visuelle et médiatique, la reconnaissance de codes télévisuels, de clins d'œil cinématographiques ou

vidéoludiques et la maîtrise de nombreuses références. Le capital culturel agit ici comme clé d'accès et la rhétorique de proximité (l'injonction aux abonnements, l'emploi de pronoms personnels tels que « nous » ou « on »), si efficace pour fédérer une communauté, peut reconduire un entre-soi.

Les signaux attentionnels identifiés sont nombreux dans la chronique, mais ils ne sont pas toujours hiérarchisés ; les incrustations textuelles, les mouvements de caméra et les variations de rythmes musicaux gagnent en efficacité lorsqu'ils orchestrent une progression argumentative explicite, quand ils s'additionnent sans priorité claire, ils ont tendance à se neutraliser. De ce point de vue, les résultats de Liang et al. (2025) sur l'humour multimodal via les danmu confirment notre intuition : la superposition de signaux peut produire tour à tour engagement ou surcharge selon la compétence de décodage et la disponibilité attentionnelle du spectateur. La même séquence peut dès lors opérer comme récompense intertextuelle pour les initiés et comme obstacle pour d'autres, ce que résume Malik dans sa remarque sur un public cherchant d'abord « à s'égayer » quelques minutes, attiré par l'attrait visuel plus que par le contenu.

Sensibiliser les publics aux procédés de manipulation visuelle et discursive n'implique pas de renoncer à la multimodalité, il s'agit au contraire d'en expliciter les mécanismes pour doter les spectateurs d'outils de vigilance. Or, nos matériaux suggèrent une ambivalence du côté des publics, entre un attrait certain pour une écriture vivante et la conscience d'une orchestration persuasive. Certains enquêtés expriment une prudence à l'égard d'une « densité » perçue comme « volonté d'imposer une lecture ». Cette lucidité peut être travaillée didactiquement, l'éducation aux médias et à l'information offrant le cadre idoine pour développer des compétences d'analyse qui rendent la multimodalité intelligible plutôt que simplement subie. Elle rejoint en cela la perspective de la littératie visuelle, non comme un savoir technique sur les images, mais bien comme le développement d'une compétence critique, culturelle et sociale de ces dernières.

Rappelons que le panel d'entretien est constitué de personnes ayant un haut potentiel cognitif et d'analyse, capable de prendre de la distance face aux logiques algorithmiques ou aux stratégies persuasives du vidéaste. Dans la chronique étudiée, toutes les démarches du vidéaste servent un projet critique explicite, lequel semble être compris par notre panel. Qu'en est-il des autres spectateurs, moins outillés ? Le risque n'est pas tant celui d'une manipulation que celui d'une sélectivité involontaire par une écriture conçue pour rendre accessibles des analyses complexes, et qui peuvent devenir, pour certains publics, un signe de fermeture. L'observation de Nicolas selon laquelle la vidéo propose « un format adressé au spectateur, presque de l'ordre de la conversation, et qui n'est pas si innocent », prend ici une portée théorique : la conversation est un dispositif, avec ses règles, ses jeux de rôle, et y participer suppose d'en reconnaître les codes…

## Limites et perspectives

Ces résultats, s'ils apportent des éléments sur la réception des vidéos politiques en ligne, doivent cependant être nuancés par les limites de l'étude. Le corpus vidéo retenu, bien que large, reste circonscrit à une seule chronique et à un seul vidéaste, ce qui limite la possibilité de généraliser les conclusions à l'ensemble des productions politiques sur YouTube. De même, l'échantillon d'entretiens, composé de neuf spectateurs, offre une richesse qualitative mais demeure très restreint. Les participants partagent souvent un intérêt préalable pour la politique ou pour le travail d'Usul, ce qui introduit de ce fait un biais de sélection et réduit la diversité des profils socioculturels représentés. Enfin, la méthodologie repose essentiellement sur l'analyse qualitative et déclarative : les propos recueillis révèlent des perceptions et des expériences, mais ne permettent pas de mesurer objectivement certains phénomènes cognitifs comme l'attention sélective, la charge mentale ou la mémorisation. Des outils expérimentaux, (eye-tracking, etc.), auraient pu compléter cette approche et offrir des données quantitatives plus robustes.

# Conclusion

La présente étude montre que la chronique Ouvrez les guillemets constitue un observatoire privilégié des écritures politiques multimodales sur YouTube. L'articulation d'images d'archives, de séquences originales, de citations savantes, de codes médiatiques et de références musicales produit une rhétorique riche pour certains, mais exigeante pour d'autres. Du côté de la production, cette hybridation s'inscrit dans une économie de remédiatisation et de remix qui confère au discours sa force argumentative et sa dimension fédératrice. Du côté de la réception, nos entretiens confirment que ces choix de production soutiennent l'intérêt, mais qu'ils sollicitent également des compétences visuelles et culturelles inégalement distribuées, générant des effets différenciés d'adhésion, de compréhension et de crédibilité.

Sur le plan théorique, le croisement entre sciences de l'information et de la communication et sciences cognitives apparaît déterminant. Les premières éclairent la logique d'intertextualité, l'ancrage sémiotique des images et la performativité des dispositifs de référence, tandis que les secondes explicitent la gestion de la charge cognitive et l'impact des émotions sur le jugement. Cette double perspective permet de comprendre pourquoi une même séquence peut, selon les profils, renforcer l'engagement ou provoquer un décrochage, et pourquoi les indices de légitimation accroissent la crédibilité chez certains tout en suscitant, chez d'autres, une méfiance face à une accumulation perçue comme ostentatoire. L'ensemble plaide pour une conception instrumentée de la multimodalité, à la fois comme une ressource persuasive

et une contrainte cognitive.

Enfin, l'étude met en avant un enjeu éducatif central. Si la multimodalité s'impose comme la grammaire ordinaire du politique en ligne, l'égalité d'accès à la compréhension des messages suppose une littératie visuelle relativement robuste. Travailler cette compétence ne revient pas à appauvrir la créativité des producteurs, mais à doter les publics d'outils d'analyse, d'indices de vigilance et de repères culturels permettant d'entrer dans le jeu de la référence sans s'y perdre, et surtout à apprécier la tentative du vidéaste. Intégrée à l'éducation aux médias et à l'information, cette littératie permettrait de transformer la puissance esthétique de l'image en ressource critique partagée, et d'envisager, face aux rhétoriques de la persuasion, un pied un temps soit peu équitable entre les spectateurs. En ce sens, la contribution la plus saillante de ce travail tient moins à la description d'un style qu'à la démonstration d'un besoin : outiller les publics pour que la richesse multimodale devienne une chance cognitive et démocratique, plutôt qu'un facteur supplémentaire d'inégalités. En outre, l'enjeu de cette littératie visuelle est de promouvoir l'exercice d'une citoyenneté éclairée et demeure une condition d'égalité face aux discours politiques qui se jouent de plus en plus de la plasticité des productions socionumériques - à voir l'utilisation des réseaux sociaux numériques par la scène politique française en temps de campagnes électorales. Au-delà de la sphère scolaire, l'intérêt de développer cette littératie visuelle est donc éminemment politique, voire militante.

# Références


Allard, L. (2009). Remix Culture : l'âge des cultures expressives et des publics remixeurs. Dans D. Pasquier, J.-P. Cottet, & F. Granjon (Éds.), Pratiques numériques des jeunes (pp. 137-146). Presses Universitaires du Septentrion.

Allard, L. (2015). La remix culture : une poïétique ordinaire du Web. In F. Kaplan & N. Nova (Eds.), La culture internet des mèmes (pp. 62-75). Presses Polytechniques et Universitaires Romandes.

Amadieu, F., & Tricot, A. (2014). Apprendre avec le numérique : mythes et réalités. Retz.

Avgerinou, M. D., & Ericson, J. (1997). A review of the concept of visual literacy. British Journal of Educational Technology, 28(4), 280–291.

Barthes, R. (1982). L'obvie et l'obtus : Essais critiques III. Seuil.

Bédard, M., Millette, M., & Brouard, F. (2022). Mèmes et culture numérique : circulation, appropriation et détournement. Revue française des sciences de l'information et de la

communication, 24.

Debes, J. (1969). The Loom of Visual Literacy. Audiovisual Instruction, 14(8), 25–27.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. Educational Psychology Review, 21(2), 113–140.

Douyère, D. (2019). Remédiatisations : circulation et transformations des images. Communication & langages, 199(1), 3–22.

Farmer, L. S. J. (2019). Visual literacy education: Utilizing digital images. Springer.

Joffe, H. (2007). Visual communication and social representation. In G. Moloney & I. Walker (Eds.), Social representations and identity: Content, process, and power (pp. 190–213). Palgrave Macmillan.

Gervereau, J.-P. (1990). Voir, comprendre, analyser les images. La Découverte

Kaufmann, J.-C. (2011). L'entretien compréhensif. Armand Colin.

Lebrun, M. (2015). La littératie visuelle comme compétence en SIC : enjeux théoriques et pédagogiques. Revue française des sciences de l'information et de la communication, 7.

Lescarret, O., Courbet, D., & Bernard, F. (2024). Emotion, trust, and persuasion in digital environments. New Media & Society. Advance online publication.

Lescarret, O., Hoareau, N., & Courbet, D. (2023). Digital persuasion: How credibility and trust shape the effects of online political communication. Journal of Computer-Mediated Communication, 28(3), 147–165.

Li, M. (2024). The influence of multimodal connectedness on political participation in China: An empirical study of the O-S-R-O-R model. Frontiers in Communication, 9.

Longhi, J. (2018). Le « genre YouTube » : émergence, codes et pratiques. Médiation et information, 47, 21–34.

Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.

Roozenbeek, J., van der Linden, S., & Nygren, T. (2022). Prebunking interventions based on the psychological inoculation theory to counter misinformation: A systematic review. Nature Human Behaviour, 6(7), 880–892.

Salojärvi, V. (2023). A rhetoric-performative multimodal analysis of the campaign videos

of two far-right Finnish parties. Journal of Scandinavian Studies in Communication, 43(4), 287–305.

Shifman, L. (2013). Memes in digital culture. MIT Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.

Törnberg, P. (2025). The aesthetics of climate misinformation. Environmental Politics.

Uminski, C. (2025). Probing visual literacy skills reveals unexpected student difficulties. CBE—Life Sciences Education, 24(1).

Zayeb, A. J., Aleidan, A. A., & Ali, N. G. (2024). The power of visuals: The significance of higher education teacher professional development in visual literacy in Kuwait. International Journal of Education and Practice, 12(1), 13–26.