

La transcription automatique comme outil d'apprentissage du FLE: quelques réflexions à partir d'un corpus oral

Alida Maria Silletti,
Università degli studi di Bari Aldo Moro
(Dipartimento di Scienze politiche)
alida.silletti@uniba.it

Abstract

This article deals with some results of a teaching experience with learners of French as a foreign language as far as the use of tools about artificial intelligence, namely automatic speech-to-text recognition, is concerned. Starting from the importance of working with ecological data, the study explores a little oral corpus dealing with Emmanuel Macron's institutional communication and its automatic transcriptions by YouTube in order to detect advantages and disadvantages of this tool. Attention has been paid to cases of phonetic identity and to omissions dealing with the results of this automatic intralinguistic translation about morphosyntax and syntactical problems linked with sentence. The last part of the article is dedicated to French oral grammar and to a particular phenomenon of syntactic interruption, namely "*parenthèses*" (Berrendonner 2008), for which in written grammar punctuation markers – which are lacking in the automatic transcriptions analysed – are essential.

Keywords: automatic speech-to-text recognition, artificial intelligence, *FLE*, syntax, teaching experience

Introduction

Les corpus s'avèrent être une ressource essentielle pour l'apprentissage de la grammaire, du vocabulaire et de spécificités du français comme langue étrangère (Auzéau, Abiad 2018) tant à l'écrit qu'à l'oral. Notre intérêt en tant qu'enseignante du français sur objectifs spécifiques (FOS) dans des filières LANSAD consiste à présenter à nos élèves¹, qui se servent de la langue française comme outil pour cerner des objectifs professionnels plutôt qu'en termes de réflexions métalangagières, une manière d'aborder des corpus oraux et surtout des phénomènes langagiers du point de vue de leur possible utilité pratique. À cet effet, nous leur proposons l'analyse de corpus oraux issus de la communication politico-institutionnelle du Président de la République française, à partir de vidéos de discours, de messages, de conférences de presse et d'entretiens qu'il a accordés à la presse nationale ou internationale depuis son élection en 2017.

La démarche d'analyse proposée à ce public inscrit en Master 1², pour qui une maîtrise de la langue française niveau B1 est demandée comme prérequis pour suivre notre cours intensif de français avancé (niveau à atteindre: B2/B2+), sur un semestre, consiste à travailler, entres autres, sur les vidéos susmentionnées à l'appui et à partir d'outils de l'intelligence

¹ L'emploi d'"élèves" au lieu d'"étudiants" ou "apprenants" s'inscrit dans notre intention d'utiliser des stratégies et des outils permettant une "neutralité discursive" (Charaudeau 2021) dans ce cas ainsi que dans tous ceux qui pourraient comporter du "masculin neutre" au sein de la présente recherche.

² Il s'agit du Master 1 en Relations internationales et études européennes et du Master 1 en Sciences des Administrations du Département de Sciences politiques de l'Université Aldo Moro de Bari.

artificielle (IA) exploités à des fins didactiques. Il s'agit de l'exercice d'observation, d'analyse et de correction en post-édition de transcriptions automatiques et de traductions automatiques du français à l'italien. Si pour cette dernière activité de traduction interlinguistique³ l'attention est focalisée sur les traductions automatiques issues des logiciels Google traducteur et DeepL, et de Reverso contexte, les transcriptions automatiques font l'objet d'un exercice de traduction intralinguistique. C'est notamment de l'observation et de la correction en post-édition de transcriptions automatiques brutes à partir des vidéos disponibles sur YouTube (en activant le sous-titrage automatique) qu'il sera question dans cette étude. Le but est de souligner la manière dont cette activité pédagogique, mobilisant l'utilisation d'outils déjà connus et employés au quotidien par les élèves, permet de travailler sur des compétences qui ne seraient pas détectées dans un cours de français langue étrangère (FLE) traditionnel. C'est en outre également pour permettre aux élèves de réfléchir sur des outils de l'IA et sur un emploi raisonné de ceux-ci que nous avons imaginé nos ateliers de transcriptions et de traductions automatiques depuis l'année universitaire 2020-2021.

Cette étude s'ouvre par l'appui sur des corpus authentiques et sur des données écologiques en classe de langue, et sur l'utilisation d'outils issus de l'IA à des fins didactiques au niveau universitaire. Nous présenterons ainsi nos ateliers pour mettre en évidence les compétences qu'une transcription automatique peut déceler en termes d'apprentissage du français. Nous montrerons des cas pratiques à partir d'un petit corpus, qui se veut un modèle pour souligner les avantages et les désavantages de l'emploi de l'IA, notamment d'une transcription automatique, en termes d'apprentissage. Notre attention portera sur des aspects généraux et sur des aspects plus ponctuels liés à l'analyse de la morphosyntaxe, de la macrosyntaxe de la phrase et de l'oral à partir des résultats issus des transcriptions automatiques. Nous proposerons une catégorisation des erreurs à partir d'une grille construite selon le degré de précision et de correction de reproduction, à l'écrit, des mots et expressions faisant l'objet de la transcription automatique. Pour ce faire, nous reviendrons sur le choix du corpus, sur les caractéristiques de cet oral non entièrement spontané et sur la manière dont il peut représenter un point de départ pour intégrer l'oral au sein d'une grammaire traditionnelle du français, basée sur l'écrit. Bien que notre étude soit surtout pratique, notre cadre théorique de référence s'appuie sur les objectifs du Cadre européen commun de référence pour les langues (CECR) pour aborder les compétences activées par l'étude de corpus oraux (Ravazzolo, Étienne 2019) en termes surtout morphosyntaxiques et de syntaxe de la phrase complexe (Blanche-Benveniste 1990; Martin 2011; Berrendonner 2008). Quant à l'emploi de l'IA à des fins didactiques, nous appuyons nos réflexions sur Raus *et al.* (2023), ainsi que sur nos recherches sur le sujet (Silletti 2022; soumis), tandis que pour l'analyse de l'outil de transcription automatique de YouTube nous nous référerons, entre autres, à un Rapport de recherche publié en 2020 (Tancoigne *et al.* 2020) comparant des outils de transcription automatique pour le français. Même si l'angle d'analyse adopté dans cette dernière étude relève surtout de l'évaluation des fonctionnalités des plateformes examinées, elle permet de mieux encadrer le logiciel de YouTube et d'en détailler les traits spécifiques.

1. Les corpus oraux comme outils didactiques d'apprentissage

Parmi les compétences pour l'apprentissage d'une langue prévues par le Cadre européen commun de référence pour les langues de 2001, actualisées par le volume complémentaire de 2020 (CECR 2021), un rôle majeur est attribué à la compétence

³ Celle-ci ne fera pas l'objet de cette étude.

interactionnelle. Le CECR précise que la compréhension orale au niveau B1 est restreinte à une langue standard ou familière et aux seules situations dans lesquelles les personnes s'expriment clairement (CECR 2021: 50), alors que le niveau B2 permet d'aboutir à une compréhension même en présence de bruits, à condition que ceux-ci ne soient pas très forts, que la structure du discours ne soit pas inadaptée ou qu'il n'y ait pas d'expressions idiomatiques (CECR 2021: 50). Plus en détail, dans le cadre des mises à jour du CECR de 2020, une section est consacrée à la compréhension audiovisuelle, à partir de documents vidéo en direct ou enregistrés, et de films. Dans le cas des niveaux B1 et B2 et de documents vidéo tels qu'une interview ou un débat télévisé, le niveau B2 permet de "repérer les points principaux d'un argumentaire ou de débats dans des programmes d'actualités et d'affaires publiques [...] en langue standard ou [dans] une variété familière" (CECR 2021: 55). C'est entre autres cette dernière compétence qui est mobilisée en classe de langue par notre atelier d'observation de transcriptions automatiques issues de documents vidéo authentiques.

L'intérêt des spécialistes à l'égard de l'emploi de documents authentiques pour l'apprentissage linguistique remonte aux années 1980 et 1990, lorsque l'attention a porté sur des enregistrements spontanés permettant d'avoir accès à des pratiques réelles de production de la parole. Il en est ainsi de l'approche du *data driven-language* combinant l'oral avec les ressources offertes par la linguistique de corpus, surtout en contexte anglophone (Aston 2001), en termes de collocations et de récurrences par le biais d'un concordancier, à partir de l'analyse des corpus oraux, donc des données de terrain transcrites automatiquement. L'observation de ces corpus révèle, comme Ravazzolo et Étienne (2019) le remarquent à partir des corpus d'interactions développés par le laboratoire ICAR de l'Université Lyon 2, de l'usage de marqueurs spécifiques de l'oral, dont la prosodie et les dislocations, et des caractéristiques qui sont liées à la langue des conversations (la superposition de parole), en plus de la présence du non verbal avec les bruits-sons. C'est également en partie ce qui émerge des travaux pionniers sur le français parlé de Claire Blanche-Benveniste (1990, entre autres) datant des années 1990 et 2000, montrant que la transcription de l'oral permet de cerner des aspects, des approches et des observables différents, négligés par la tradition grammaticale fondée sur la langue écrite, qui sont à la base de la construction d'une grammaire de l'oral. Nous reviendrons sur cet aspect au paragraphe 5.2, lorsque nous nous intéresserons au phénomène morphosyntaxique des "parenthèses" (Berrendonner 2008).

2. Le corpus: critères de sélection et d'analyse

Si, en didactique des langues, l'emploi de larges collections d'exemples⁴ tirés de données écologiques et sans aucune intervention exogène vise surtout à faire familiariser les élèves au français parlé en interaction dans des situations sociales variées, l'oral sur lequel nous appuyons cette recherche est partiellement contrôlé tant en termes de sujets locuteurs que de thèmes abordés. Il s'agit de documents audiovisuels dont la transcription automatique sert de modèle représentatif de corpus oraux plus vastes disponibles sur YouTube. Ces vidéos relèvent de la communication politico-institutionnelle du Président de la République française, Emmanuel Macron, au mois d'août 2023, et de leur transcription générée automatiquement à partir du sous-titrage automatique, gratuit, de YouTube pour le français. Pour cette étude, nous nous intéresserons à des cas de communication à une seule voix. Cet oral monogéré, non entièrement spontané, est encadré au sein d'un discours institutionnel (Oger 2005), où il est possible de retrouver "des fonctionnements institutionnels, des activités

⁴ Rappelons, à ce propos, les travaux de C. Blanche-Benveniste (1990) et de M. Avanzi (2011).

routinisées ou ritualisées où [les discours institutionnels] prennent sens” (Oger 2005: 113). Cet oral, qu’E. Macron a préparé au préalable et qu’il peut même lire, est encadré au sein d’un “discours d’autorité” (Krieg-Planque 2015: 115). Il est linéaire et apte à receler des traits de stabilité, et il est accompagné d’une transcription automatique.

La situation d’énonciation (Maingueneau 2004) voit ainsi un système abstrait composé d’un sujet énonciateur et locuteur face à des sujets *in absentia* – dans le cas des messages envoyés devant une caméra et à l’Élysée – ou *in praesentia*, comme il arrive aux discours du Président de la République. Ces derniers sont produits face à un public réel et à dans un lieu qui acquiert un caractère institutionnel. La “scénographie” (Maingueneau 2004) y est donc “canonique” et relève du genre de la déclaration du Président de la République tant dans le cas des messages que des discours présidentiels.

Notre intérêt pour ces vidéos porte sur la manière dont le logiciel de transcription automatique de YouTube reproduit, à l’écrit, ce qui est énoncé, que nous essaierons de détailler au paragraphe suivant.

3. L’IA à des fins didactiques: le logiciel de transcription automatique de YouTube

Depuis les années 1980 et 1990, lorsque les outils d’enregistrement de la parole se démocratisent, le recueil de la parole acquiert un rôle de plus en plus central au sein de domaines différents des sciences humaines et sociales. Des difficultés sont pourtant liées surtout à des temps de production étendus et à des résultats parfois insatisfaisants en termes qualitatifs, mais l’intégration de modules d’IA aux algorithmes de reconnaissance automatique de la parole permet à ces outils d’améliorer leur performance, d’où une activité automatique de transcription beaucoup plus satisfaisante (Tancoigne *et al.* 2020).

Les outils d’enregistrement de la parole transformée en texte par le biais de sous-titres, monolingues ou traduits, ont débuté par la traduction automatique interlinguistique pour ensuite concerner également la traduction intralinguistique ou monolingue. Le résultat en est que la production automatique de sous-titres monolingues ou traduits *via* des modèles neuronaux d’apprentissage profond est devenue un champ d’application fécond du traitement automatique des langues (Buet, Yvon 2021).

Un rapport de recherche novateur, publié en 2020 dans le cadre d’un vaste partenariat scientifique (Tancoigne *et al.* 2020), s’intéresse à la comparaison entre huit outils de transcription automatique⁵ à partir de tests conduits sur des extraits de fichiers audio en langue française produits dans des situations différenciées, tant monologiques que dialogales. Parmi les outils testés, il y a également celui de YouTube, qui est examiné, de même que les autres plateformes, en termes de fonctionnalités et d’évaluation des transcriptions obtenues. Les résultats de ce rapport – auquel nous renvoyons pour plus de précisions également sur la méthodologie utilisée – montrent que le logiciel utilisé par YouTube ne figure pas parmi les plus performants mais que ses coûts sont modérés (Tancoigne *et al.* 2020).

Ce sont surtout la disponibilité gratuite en ligne et la possibilité d’effectuer des transcriptions automatiques à partir de vidéos dont la durée est étendue qui nous ont fait pencher, en 2020, à peu près lorsque ce rapport est publié, pour l’outil de transcription automatique de YouTube dans le cadre de l’atelier de notre cours de français avancé⁶. Il

⁵ Il s’agit de Go Transcribe, Happy Scribe, Headliner, Sonix, Video Indexer, Vocalmatic, Vocapia et YouTube.

⁶ L’idée de créer ces ateliers nous a été inspirée par notre participation au volet didactique du projet européen *Artificial Intelligence for European Integration* (AI4EI), dirigé par Rachele Raus, financé par le Centre d’excellence Jean Monnet de l’Université de Turin, <https://www.jmcoe.unito.it/content/kick-conference-ai4ei>.

s'agissait en outre d'un outil que nos élèves connaissaient déjà, qu'il nous semblait intéressant de leur faire utiliser pour un emploi plus réfléchi. Le troisième critère de choix du logiciel de YouTube a concerné les résultats des transcriptions automatiques produites. Bien que les performances de cet outil s'améliorent sous certaines conditions, des problèmes importants – qui représentent, à l'inverse, le point de force de nos activités en classe de langue – sont toujours présents, lesquels feront l'objet de notre analyse du corpus.

4. Le projet didactique pilote sur les transcriptions automatiques: premières applications

L'atelier d'observation et de correction des transcriptions automatiques est adressé à des élèves ayant obtenu leur licence tant en Sciences politiques ou administratives qu'en Sciences de la communication, en Histoire, en Travail social, voire en LEA, en Traduction spécialisée ou en Langues et littératures étrangères. D'où des compétences très différenciées en FLE, du B1 au B2, voire au C1. De surcroît, le M1 en Relations internationales et études européennes voit également la présence de quelques élèves provenant d'un contexte international lié aux migrations internationales et à l'asile, dont des pays d'Afrique noire francophone. Or, malgré l'hétérogénéité de ce public, il lui est demandé de posséder un niveau de départ de connaissance de la langue française correspondant au moins à B1 (pour aboutir à un niveau B2/B2+). En outre, rappelons qu'aucun cours de lectorat avec du personnel enseignant de langue maternelle ne lui est dispensé. Ces élèves ne suivent ainsi que le cours magistral intensif de 64 heures (sur un semestre) réparti sur douze semaines, au sein duquel diverses compétences et divers objectifs doivent être mobilisés. Il s'agit de connaissances tant grammaticales, appliquées aux documents audiovisuels, y compris des aspects relevant du français parlé et d'une grammaire du français oral, que centrées sur des objectifs spécifiques des Masters concernés, notamment la géopolitique et la diplomatie internationale. Les ateliers sur l'IA correspondent à environ 30 heures de cours au total.

Plutôt qu'un travail sur des extraits de documents audiovisuels, nous proposons une réflexion sur des documents authentiques considérés dans leur intégralité. Diverses raisons sont à la base de ce choix: l'appui sur le contexte situationnel; la compréhension globale du thème abordé; la possibilité, pour l'élève, d'avoir une première compréhension générale de tout le document à exploiter. C'est pourquoi les documents audiovisuels retenus portent sur des sujets déjà connus par les élèves en raison des thèmes de politique internationale traités dans d'autres cours de leur Master ou des connaissances préalables acquises à partir du cours magistral de français avancé.

La première étape de l'atelier de transcription automatique concerne l'écoute globale de la vidéo, sans sous-titres mais après en avoir examiné, dans les grandes lignes, le titre et la date de réalisation et parution. Ces informations sont disponibles sur YouTube avant d'activer la vidéo. Cette phase d'entrée dans les données est suivie par une deuxième visualisation de la vidéo avec les sous-titres générés automatiquement depuis YouTube en activant la fonction "sous-titres", en bas de l'encadré de la vidéo, ou "transcription". Par cette dernière option, un encadré s'ouvre à droite de la vidéo avec la transcription, pourvue de l'édition des balises de temps – ce qui est le cas pour toute transcription automatique depuis YouTube –, qui s'étend au fur et à mesure que la vidéo se poursuit en colorant en gris ou en bleu clair la partie qui est prononcée (Fig. 1).

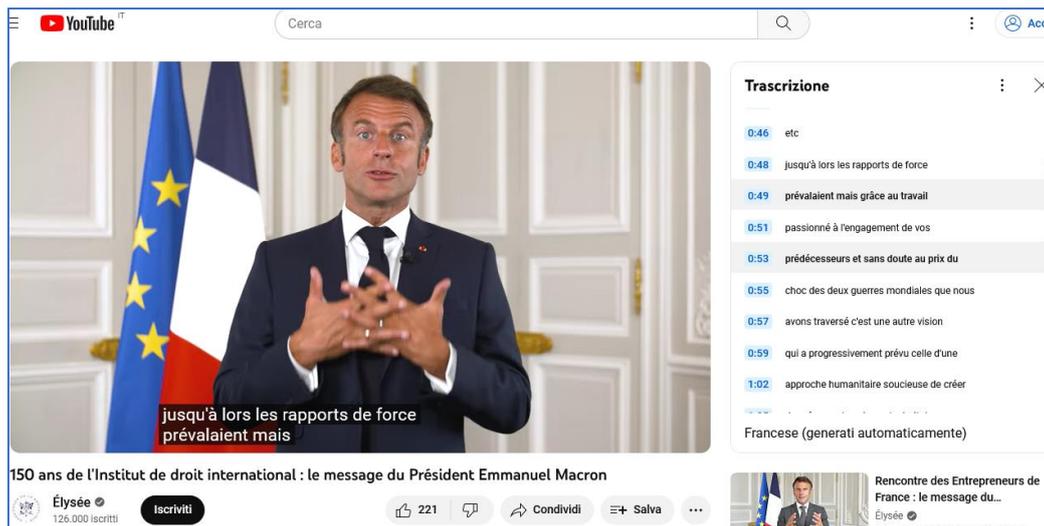


Figure 1. Exemple de transcription automatique de YouTube (<https://www.youtube.com/watch?v=K3DBYafYpHo>).

Dans la troisième phase, les élèves se trouvent confrontés à l'exercice de réflexion et de correction de la transcription automatique. Celui-ci est réalisé à partir d'un fichier word préalablement préparé par nous-même, comportant un tableau à deux colonnes: celle de gauche contient la transcription automatique brute, tandis que celle de droite devra contenir la transcription révisée à partir de la version automatique fournie par l'outil de transcription automatique de YouTube. L'exercice demandé relève donc d'une traduction intralinguistique en parallèle, avec la transcription automatique à gauche et la colonne de droite à remplir avec la transcription révisée effectuée par chaque élève en post-édition.

Il est évident que le degré de précision d'une transcription automatique brute est meilleur si le document audiovisuel est dépourvu de bruits de fond et qu'il est doté d'une qualité de captation optimale (Tancoigne *et al.* 2020), mais aussi s'il s'agit d'une communication gérée par un seul sujet et avec une seule voix, autrement dit dépourvue de tout ce qui relève d'une communication polygérée, comme une interview ou une conférence de presse.

En tout état de cause, même si une précision absolue du discours dans les transcriptions automatiques de YouTube n'a pas lieu, il est possible de relever une reconnaissance globale du sens sous les conditions de production que nous avons annoncées. Cependant, pour ce faire, il faut que les règles orthographiques, de syntaxe et d'accord soient respectées – ce qui n'est pas toujours le cas, comme nous le montrerons au paragraphe suivant.

5. La détection et la catégorisation des erreurs

Le petit corpus que nous exploitons pour la présente recherche est composé d'un message du Président de la République prononcé à l'occasion du 150^{ème} anniversaire de l'Institut de droit international, à Angers, le 27 août 2023 (transcription automatique TA1), et du discours du Président de la République lors de la Conférence annuelle des Ambassadrices et des Ambassadeurs, à Paris, le 28 août 2023 (TA2). Ces deux documents audiovisuels sont similaires à divers égards, même si leur longueur et le nombre de mots y diffèrent – 1h54 minutes d'enregistrement pour le discours, 5min36 secondes pour le message. Nous avons traité à des genres de discours monologiques issus de la communication politico-institutionnelle

du Président de la République, dont les mots sont bien scandés et le débit est lent. Si l'on mesurait le degré de complexité de ces documents audiovisuels par rapport aux conférences de presse et aux interviews du Président de la République, ils figureraient sur un niveau de difficulté bas. Enfin, la connaissance, en amont, de la date et heure, et du lieu de prononciation de ces déclarations contribuent à rendre ces contenus "stables" en termes d'exploitation et de traitement automatique – d'où une bonne qualité audio. Cela confirme l'intérêt de nous y appuyer en tant qu'entraînement pour nos élèves, à savoir avant d'aborder des documents audiovisuels plus complexes.

Afin de proposer ces activités d'observation et de correction collectives de transcriptions automatiques, il faut des vérifications préalables. Notre évaluation des transcriptions est basée sur les transcriptions automatiques proposées par YouTube et non pas sur une retranscription manuelle. Cela permet non seulement de mesurer, en amont, la complexité du document transcrit automatiquement et les problèmes orthographiques et morphosyntaxiques figurant dans la transcription automatique, mais surtout d'effectuer la même activité et les mêmes démarches qui sont par la suite proposées à nos élèves.

Le degré de correction des deux vidéos et leurs transcriptions automatiques sont assez bons en raison des genres de discours choisis, de la qualité des vidéos – qui ne comportent ni bruits de fond ni (d'importants) problèmes de captation des mots –, du débit du sujet locuteur – qui est entre autres à relier à son rôle et à sa posture. C'est à partir d'un classement des problèmes relevés en termes d'abord de perception auditive que nous avons créé une grille des erreurs relevées⁷ en distinguant des erreurs morphosyntaxiques et syntactico-phrastiques comportant une substitution avec identité phonétique de ceux qui portent sur une omission – inscrite, encore une fois, dans la morphosyntaxe ou dans la syntaxe de la phrase⁸.

5.1 Substitution avec identité phonétique

En termes morphosyntaxiques et syntactico-phrastiques, nous avons relevé des erreurs récurrentes correspondant à des types divers, liées à des coïncidences phonétiques avec substitution d'un mot plein – un verbe conjugué ou un participe passé – par un autre mot plein appartenant à la catégorie du verbe. Ces cas s'avèrent être nombreux et dépendent, à notre avis, de deux situations phrastiques principales.

La première, illustrée par l'ex. 1, est caractérisée par un accord erroné entre le participe passé d'un temps verbal composé – voix active ou passive – et le sujet syntaxique de la phrase, auquel le participe passé doit se référer, en particulier si ce sujet est représenté par le pronom relatif "qui". Ces problèmes s'accroissent surtout si le sujet syntaxique est séparé du verbe par des mots, quelle que soit leur catégorisation morphologique: il peut s'agir d'adverbes, de locutions adverbiales ou prépositionnelles. Tel est également le cas du participe passé non ou mal accordé, ou bien du participe passé référé au pronom relatif complément d'objet direct "que" (ex. 2):

⁷ Puisque notre recherche se veut qualitative, nous n'allons pas présenter des chiffres comportant la fréquence de telle ou telle erreur – dont le calcul statistique serait *a priori* biaisé en raison de la taille du corpus.

⁸ Des cas de substitution avec et sans proximité phonétique sont également présents dans le corpus mais ceux-ci relèvent d'aspects sémantiques, morphologiques et lexicaux, qui ne font pas l'objet de la présente étude.

Ex. 1 (TA1)⁹

[...] internationale cette 2:42 réponse nous l'avons porté au nom du 2:43 droit international et de ses valeurs et 2:45 ce avec détermination et je le redis ici 2:46 avec beaucoup de clarté nous avons pris 2:48 plusieurs résolutions très claires qui 2:51 ont été votés par l'Assemblée générale 2:52 des Nations unies à une [...]	[...] internationale. Cette 2:42 réponse, nous l'avons portée au nom du 2:43 droit international et de ses valeurs, et 2:45 ce avec détermination et – je le redis ici 2:46 avec beaucoup de clarté – nous avons pris 2:48 plusieurs résolutions très claires qui 2:51 ont été <u>votées</u> par l'Assemblée générale 2:52 des Nations unies à une [...]
--	--

Ex. 2 (TA1)

mais grâce au travail 0:51 passionné à l' engagement de vos 0:53 prédécesseurs et sans doute au prix du 0:55 choc des deux guerres mondiales que nous 0:57 avons traversé c'est une autre vision 0:59 qui a progressivement prévu celle d'une 1:02 approche humanitaire	[...] mais grâce au travail 0:51 passionné, à l'engagement de vos 0:53 prédécesseurs et sans doute au prix du 0:55 choc des deux guerres mondiales que nous 0:57 avons <u>traversées</u> , c'est une autre vision 0:59 qui a progressivement prévalu, celle d'une 1:02 approche humanitaire [...]
--	---

Ces problèmes d'accord du participe passé dépendant de pronoms relatifs simples peuvent être dus au fait que le logiciel de transcription automatique n'est pas en mesure d'identifier l'antécédant du pronom relatif simple concerné. Ce dernier, qui est invariable, est épïcène mais aussi dépourvu de nombre.

La seconde situation d'identité phonétique engendrant un accord erroné du participe passé d'un temps verbal composé (ex. 3) montre un aspect intéressant, que nous avons relevé dans d'autres exemples du corpus: le logiciel ne réussit pas à associer le son [z] à une liaison entre un participe passé se terminant par “-s” et le mot qui suit se commençant par une voyelle, lui préférant un accord au féminin, non pertinent:

Ex. 3 (TA2)

zacharũlo chacun d'entre vous 1:00 s'est mise au service des Français je vous en remercie [...]	[...] Zacharopoulou. Chacun d'entre vous s'est <u>m̃is</u> au service des Français. Je vous en remercie. [...]
--	--

⁹ Dans les exemples, la colonne de gauche présente la transcription automatique brute de YouTube, alors que la colonne de droite contient nos corrections en post-édition, identifiées par le soulignage, par rapport aux aspects analysés.

5.2 Omission

Pour ce qui concerne les omissions au niveau morphosyntaxique ou phrastique, celles-ci peuvent être réparties en deux sous-catégories. La première porte sur le manque de mots ou expressions comportant des retombées syntaxiques (ex. 4). Le logiciel de YouTube ne perçoit pas, dans cet extrait, deux monosyllabes réellement prononcés, générant également des problèmes sémantiques de compréhension de la phrase – il s’agit du présent de l’indicatif du verbe “avoir”, “a”, et du pronom adverbial “y” se combinant avec *compris*:

Ex. 4 (TA2)

<p>[...] il faut bien le dire l'Occident avait une place prépondérante et encore une 4:10 place prépondérante avec d'abord la montée de la guerre 4:15 compris sur le sol européen j'y reviendrai dans un instant la politique aussi de l'état de fête qui des confins [...]</p>	<p>Il faut bien le dire: l'Occident avait une place prépondérante et <u>a</u> encore une 4:10 place prépondérante, avec d'abord la montée de la guerre, 4:15 <u>y</u> compris sur le sol européen – j'y reviendrai dans un instant –, la politique aussi de l'état de fait qui, [...]</p>
--	---

Le résultat du logiciel, comme les autres qui comportent la non reproduction de monosyllabes, est à rapporter à la situation d’énonciation, puisque le locuteur augmente son débit et diminue le ton de la voix dans le cas de “y” dans *y compris*. En revanche, l’occurrence qui suit immédiatement après est détectée par le logiciel en raison d’une énonciation plus nette.

C’est pourtant la seconde sous-catégorie d’omissions syntaxiques qui est la plus répandue étant donné sa fréquence élevée. Il s’agit de l’absence de signes graphiques et surtout de la ponctuation, laquelle fait totalement défaut dans les transcriptions automatiques de YouTube. Ces omissions sont par ailleurs aussi bien les plus visibles même de prime abord que celles qui affectent davantage la compréhension d’un segment, voire du document dans son intégralité. Les cas de signes graphiques omis les plus évidents portent sur l’emploi des majuscules dans les termes d’adresse – que nous n’étudions pas dans cet article – ainsi que, surtout en début de phrase, pour signaler la convention graphique de commencer une phrase écrite par une initiale majuscule après un signe de ponctuation fort (le point final, le point d’exclamation, le point d’interrogation). D’où l’absence tant de ces signes que d’une ponctuation faible (virgules) ou d’une ponctuation d’encadrement (Le Goffic 2005), permettant d’isoler un segment phrastique par des virgules doubles, de tirets, voire des parenthèses. L’exercice de révision des transcriptions automatiques est ainsi essentiel pour souligner l’importance de la ponctuation et, donc, de la syntaxe de la phrase complexe, ce qui demande, entre autres, de tenir compte d’une grammaire associant l’écrit et l’oral au niveau de la phrase. C’est pourquoi la définition même de phrase doit intégrer une dimension énonciative par rapport à sa conception, traditionnelle, dans le cadre de l’écrit.

À cet égard, nous nous inscrivons dans l’approche de P. Le Goffic (2005: 57), qui définit la phrase une totalité structurante mais aussi un acte en tant qu’énoncé unique inséparable de son énonciation, parce qu’elle permet d’étudier, entre autres, la ponctuation et, à son intérieur, des aspects tant micro- que macrosyntaxiques. Tel est le cas des éléments non régis qui ne relèvent pas strictement de la rection par rapport à un prédicat, mais surtout, en termes de ponctuation, du rôle de la rection dans l’intrication de phrases. À cet égard, deux phrases peuvent porter sur deux niveaux énonciatifs différents sans cependant s’y confondre, ce qui est marqué par l’intonation à l’oral et par la ponctuation à l’écrit. Pour étudier l’un de

ces phénomènes syntaxiques¹⁰, nous nous pencherons sur ce qu’A. Berrendonner (2008) qualifie de “parenthèses” – illustrées à partir de l’ex. 1 (“je le redis ici avec beaucoup de clarté”) et de l’ex. 4 ci-dessus (“j’y reviendrai dans un instant”). Cet auteur les considère des routines macrosyntaxiques par lesquelles le sujet locuteur interrompt de façon momentanée une unité communicative pour qu’une autre unité communicative soit entre-temps accomplie. En effet, dans les exemples, E. Macron suspend son énonciation principale le temps d’en accomplir une autre, secondaire et dépourvue de liens syntaxiques avec la première, laquelle est reprise immédiatement après. Ces deux “parenthèses” sont “canoniques” tant en termes de configuration textuelle médiane – les “parenthèses” sont mobiles dans la phrase, pouvant occuper une position même initiale ou finale (Berrendonner 2008) –, que de paramètres constitutifs de celles-ci. À ce propos, A. Berrendonner (2008: 7) rappelle qu’elles doivent comporter une unité communicative inachevée (A1); un achèvement, par la suite, de la première unité (A2); entre les deux, une “insertion parenthétique” comme séquence intercalaire exogène. Comme A. Berrendonner (2008: 8) le signale, en termes de processus discursif, une “interruption parenthétique” est une “conduite communicativement marquée”. Par exemple, E. Macron s’en sert pour renforcer son dire et son action, comme en témoigne, entre autres, l’emploi de la 1^{ère} personne du singulier. Or, si une insertion, telle qu’une “parenthèse” au sein d’une unité communicative, est tout naturellement marquée par la prosodie à l’oral, tel n’est pas le cas à l’écrit, lorsque des signes de ponctuation doivent mettre en évidence, même en termes graphiques, cette interruption. Dans une “parenthèse” médiane, ceux-ci apparaissent sous forme double *via* des virgules doubles, des tirets doubles, voire des parenthèses graphiques. À notre avis¹¹, les tirets doubles sont souvent les signes les plus aptes à indiquer des “parenthèses” du fait de leur séparation, même spatiale, plus forte que des virgules mais plus faible que des parenthèses graphiques. Ces dernières tendent à signaler des informations sémantiquement accessoires – ce qui n’est pas le cas pour les “parenthèses” syntaxiques dans le discours d’un orateur.

D’autres aspects syntaxiques et relevant d’une grammaire de l’oral mériteraient d’être explorés, parmi lesquels des dislocations à droite ou à gauche¹². Ces aspects, qui n’ont pas pu faire l’objet de la présente étude, montrent quand même que les transcriptions automatiques représentent un champ d’analyse fécond pour détecter des phénomènes syntaxiques, entre autres, de la grammaire de l’oral.

6. Conclusions

Deux aspects essentiels émergent du travail mené à l’égard des transcriptions automatiques examinées dans la présente étude, qu’il est possible de généraliser à l’IA utilisée à des fins didactiques¹³. D’une part, l’appui sur une transcription automatique permet, en soi, une épargne de temps considérable en termes de reproduction, à l’écrit, de l’oral relevant d’une vidéo – bien que le résultat soit à vérifier et à corriger; de l’autre, les transcriptions automatiques de YouTube ne sont pas optimales. Le problème principal qui est apparu à partir des vidéos et des extraits examinés, c’est la non prise en compte du contexte phrastique et du

¹⁰ L’intrication de phrases devrait prendre en compte également la prosodie. En raison de contraintes d’espace, nous renvoyons à Martin (2011) ainsi qu’à Silletti (soumis) pour l’identification des “parenthèses” à l’appui, entre autres, de la prosodie.

¹¹ Nous tenons à préciser que nos transcriptions révisées ne constituent qu’une version parmi des solutions de transcription possibles.

¹² Tel est le cas, entre autres, de “Cette réponse, nous l’avons portée [...]” (ex. 1).

¹³ Nous renvoyons, à ce propos, à Raus *et al.* (2023).

contexte situationnel de la part du logiciel de transcription automatique. Cela est prouvé par la non identification des liaisons et des accords de participes passés dépendant de pronoms relatifs simples, ainsi que, sous un autre aspect, par les noms communs de fonction dépourvus d'initiales majuscules. En outre, si le degré de précision le plus élevé est atteint par rapport au lexique, les cas d'homophonie interviennent parfois sur le sens de la phrase, même en le détournant. C'est pour autant aux niveaux morphosyntaxique et surtout syntactico-phrastique que la transcription automatique se révèle plus déficitaire à cause de l'absence de signes de ponctuation et en partie de signes graphiques. D'où l'intérêt de nous pencher sur ce dernier aspect pour réfléchir sur l'importance de la ponctuation et, en amont, sur la nécessité d'intégrer une grammaire de l'oral à la tendance grammaticale traditionnelle, fondée sur l'écrit. Cela permettrait, entre autres, de faire référence à la prosodie et de cerner des phénomènes macro- et microsyntaxiques tels que les "parenthèses" ou les dislocations. Il s'avère, en effet, que ces structures sont fréquemment utilisées par les sujets locuteurs et orateurs pour adapter leurs discours à l'auditoire en vue de le persuader, en insistant sur les aspects qu'ils considèrent plus saillants.

En dépit des remarques liées à l'outil de transcription automatique de YouTube, l'appui sur la transcription automatique et les logiciels qui cherchent de plus en plus à la perfectionner soulignent que des frontières restent encore à franchir. Parmi celles-ci, rappelons la possibilité d'avoir affaire à de bonnes transcriptions en termes qualitatifs au-delà du type de fichier soumis en entrée. Sous cet aspect, tant le rapport de Tancoigne *et al.* (2020) que notre étude montrent que tel n'est pas le cas à l'heure actuelle. D'où l'impératif d'une vérification humaine, en post-édition, qui n'excède quand même pas le temps qu'un sujet humain prendrait s'il effectuait une transcription entièrement manuelle (Tancoigne *et al.* 2020). Cela souligne donc que le gain de temps continue à rester une variable importante lorsqu'il s'agit de mesurer les avantages et les désavantages de l'IA, même à des fins didactiques.

Enfin, l'intérêt de notre public étudiant pour les ateliers d'IA que nous avons conduits et, de notre côté, les résultats intéressants qui émergent de leurs réflexions sur les transcriptions et traductions automatiques – deux épreuves leur sont à ce propos soumises en fin de cours – nous ont amenée à poursuivre cet atelier même pendant l'année universitaire 2023-2024.

Références bibliographiques¹⁴

- Aston G. (dir.), 2001, *Learning with Corpora*, Bologne, Clueb editrice.
- Auzéau F., Abad L., 2018, "Le corpus: un outil inductif pour l'enseignement-apprentissage de la grammaire", in *Synergies France*, 12, pp. 175-187.
- Avanzi M., 2011, "La dislocation à gauche avec reprise anaphorique en français parlé. Étude prosodique", Actes de IDP09, http://unine.ch/files/live/sites/structuration_periodes/files/shared/new_am/35_.pdf
- Berrendonner A., 2008, "Pour une praxéologie des parenthèses", in *Verbum*, XXX/1, pp. 5-23.
- Blanche-Benveniste C., 1989, "Constructions verbales « en incisives » et rection faible des verbes", in *Recherches sur le français parlé*, 9, pp. 53-73.
- Blanche-Benveniste C., 1990, *Le français parlé – études grammaticales*, Parigi, Éditions du CNRS.

¹⁴ Les liens hypertexte ont été vérifiés à la date du 7 septembre 2023.

- Buet F., Yvon F., 2021, “Vers la production automatique de sous-titres adaptés à l’affichage”, in *Actes de la 28^e Conférence sur le Traitement Automatique des Langues Naturelles*, vol. 1, conférence principale, Lille, ATALA, pp. 91-104.
- Conseil de l’Europe, 2021, *Cadre européen commun de référence pour les langues: apprendre, Enseigner, évaluer. Volume complémentaire*, Strasbourg, Éditions du Conseil de l’Europe, Strasbourg, www.coe.int/lang-cecr
- Charaudeau P., 2021, *La langue n’est pas sexiste. D’une intelligence du discours de féminisation*, Lormont, Le Bord de l’eau.
- Krieg-Planque A., 2015, “Construire et déconstruire l’autorité en discours. Le figement discursif et sa subversion”, in *Mots. Les langages du politique*, 107, pp. 115-132.
- Le Goffic P., 2005, “La phrase revisitée”, in *Le français aujourd’hui*, 148/1, pp. 55-64.
- Maingueneau D., 2004, *Le discours littéraire: Paratopie et scène d’énonciation*, Paris, Armand Colin.
- Martin P., 2011, “Ponctuation et structure prosodique”, in *Langue française*, 172, 4, pp. 99-114.
- Oger C., 2005, “L’analyse du discours institutionnel entre formations discursives et problématiques socio-anthropologiques”, in *Langage et société*, 114/4, pp. 113-128.
- Raus R., Humbley J., Silletti A.M., Zollo S.D. (dir.), 2023, “Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence”, numéro special, *De Europa*.
- Ravazzolo E., Étienne C., 2019, “Nouvelles ressources pour le FLE à partir des études en interaction”, in *Linx* [En ligne], 79, document 2, mis en ligne le 30 décembre 2019, <http://journals.openedition.org/linx/3454>
- Silletti A.M., 2022, “La transcription automatique, outil pédagogique pour les étudiant(e)s en sciences politiques: quelques réflexions sur la syntaxe”, in Concetta Cavallini, Marco Santipolo (dir.), *Educare alle lingue straniere: Frontiere interdisciplinari teoriche, metodologiche e operative*, Bari, Cacucci Editore, Collana Sguardi sulla Modernità, n. 2, pp. 101-118.
- Silletti A.M. (soumis), “Les «parenthèses» à l’épreuve de la transcription générée automatiquement: une analyse syntactico-prosodique et argumentative”.
- Tancoigne E., Corbellini J.-P., Deletraz G., Gayraud L., Ollinger S., *et al.*, 2020, “La transcription automatique: un rêve enfin accessible? Analyse et comparaison d’outils pour les SHS. Nouvelle méthodologie et résultats”, *Rapport de recherche, MATE-SHS*, halshs-02917916v2