



**HAL**  
open science

## Étude des facteurs affectant la compréhensibilité de documents multimodaux : une étude expérimentale

Estelle Randria, Lionel Fontan, Maxime Le Coz, Isabelle Ferrané, Julien Pinquier

### ► To cite this version:

Estelle Randria, Lionel Fontan, Maxime Le Coz, Isabelle Ferrané, Julien Pinquier. Étude des facteurs affectant la compréhensibilité de documents multimodaux : une étude expérimentale. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 1 : Journées d'Études sur la Parole, 2020, Nancy, France. pp.534-542. hal-02798583v3

**HAL Id: hal-02798583**

**<https://hal.science/hal-02798583v3>**

Submitted on 23 Jun 2020

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

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

# Étude des facteurs affectant la compréhensibilité de documents multimodaux : une étude expérimentale

Estelle Randria<sup>1,2</sup>, Lionel Fontan<sup>2</sup>, Maxime Le Coz<sup>2</sup>, Isabelle Ferrané<sup>1</sup>, Julien Pinquier<sup>1</sup>

(1) IRIT-UPS, 118 route de Narbonne 31062 Toulouse, France

(2) Archean LABS, 20 place Prax-Paris, 82000 Montauban, France

{estelle.randria, isabelle.ferrane, julien.pinquier}@irit.fr, {lfontan, mlecoz}@archean.tech

## RÉSUMÉ

---

La compréhensibilité de documents audiovisuels peut dépendre de facteurs propres à l'auditeur/spectateur (ex. langue maternelle, performances cognitives) et de facteurs propres aux contenus des documents (ex. complexité linguistique, intelligibilité de la parole). Dans ces travaux, nous étudions les effets de facteurs propres aux contenus sur la compréhensibilité de 55 dialogues extraits de films, présentés à 15 experts (enseignants de français langue étrangère) selon cinq modalités différentes (transcription, transcription + audio, audio, audio + vidéo, transcription + audio + vidéo). Les experts ont évalué les dialogues en termes de compréhensibilité générale, de complexité du vocabulaire, de complexité grammaticale, et d'intelligibilité de la parole. L'analyse de leurs évaluations montre que (1) la complexité du vocabulaire, la complexité grammaticale, et l'intelligibilité de la parole sont significativement corrélées à la compréhensibilité générale, et (2) que les évaluations de compréhensibilité générale ont tendance à être plus élevées lors de présentations multimodales.

## ABSTRACT

---

### **Factors affecting the comprehensibility of multimodal documents : an experimental study**

It has been shown that the comprehensibility of audiovisual documents can be influenced both by factors related to the recipient (e.g., native language, cognitive performance) and factors related to the document itself (e.g., linguistic complexity, speech intelligibility). The aim of this study was to investigate the effects of the document-related factors on the comprehensibility of 55 movie clips, presented to 15 experts (language teachers) under five different (combinations of) presentation modalities (text transcript, audio, text transcript + audio, audio + video, text transcript + audio + video). The experts rated the documents in terms of overall comprehensibility, lexical complexity, grammatical complexity, and speech intelligibility. The analysis of the experts ratings indicated that (1) lexical complexity, grammatical complexity, and speech intelligibility are all three significantly correlated to the ratings of comprehensibility and (2) that the combination of different presentation modalities lead to higher ratings of comprehensibility.

**MOTS-CLÉS :** corpus, film, compréhensibilité, compréhension orale, multimodalité, complexité linguistique, complexité grammaticale, intelligibilité de la parole.

**KEYWORDS:** corpus, movie, comprehensibility, listening comprehension, multimodality, linguistic complexity, grammatical complexity, speech intelligibility.

---

# 1 Introduction

Le développement et la popularisation des plateformes de vidéo à la demande a eu un impact majeur sur la diffusion de contenus audiovisuels. La compréhension de contenu est un élément clef pour favoriser l'accessibilité à ces collections massives de documents audiovisuels. Il est possible de classer du contenu audiovisuel par genre, et il existe également des labels permettant de définir l'âge minimum conseillé pour regarder une vidéo, mais aujourd'hui il n'existe pas de classification permettant d'indiquer à quel point un document audiovisuel est compréhensible pour un public donné. En effet, ce type de document peut être perçu comme plus ou moins compréhensible selon la personne qui y a accès (personnes âgées, enfants, étrangers...). Dans ce contexte, il est intéressant d'étudier si ce type de classification est envisageable. Cet article traite de la compréhensibilité de contenu audiovisuel, plus spécifiquement de la difficulté potentielle de compréhension qui peut leur être associée, qui sera appelée par la suite « difficulté globale ». Dans un premier temps, la difficulté globale sera abordée principalement dans le cadre de l'apprentissage des langues. Les objectifs de notre étude sont : (1) d'étudier de quelle façon les humains perçoivent la difficulté globale de contenus audiovisuels (2) de savoir ce qui influence cette difficulté entre la complexité du vocabulaire, la complexité grammaticale et l'intelligibilité de la parole (c'est-à-dire à quel point la parole est facile à percevoir) et puis (3) de savoir si l'accès aux différentes modalités (vidéo, audio et texte) prises séparément ou combinées, joue un rôle sur la perception humaine de la difficulté globale.

## 2 Etudes autour de la compréhensibilité de documents

La compréhensibilité de documents, que ce soit de documents écrits, audio ou audiovisuels, a fait l'objet de nombreuses études. La grande majorité a cependant porté sur la compréhensibilité de textes et de documents audio plutôt que sur les contenus audiovisuels. Dans cette partie, nous présentons les éléments les plus souvent dégagés par les études comme étant des facteurs de complexification ou de facilitation de la compréhension orale.

Tout d'abord, il est important d'aborder des aspects linguistiques : il a en effet été démontré très tôt que le vocabulaire et la grammaire (notamment la syntaxe), jouaient un rôle prépondérant dans la compréhension écrite et orale. Les premières preuves ont été apportées dans des études portant sur la lisibilité (*readability*), qui prennent en compte les aspects linguistiques et conceptuels d'un texte pour évaluer sa difficulté globale. De nombreuses formules de lisibilité, prédisant la difficulté des textes, font appel à des descripteurs liés à la difficulté du vocabulaire et à la complexité grammaticale. Les formules les plus connues sont celles de Lively et Pressey ([Lively & Pressey, 1923](#)), Flesch ([Flesch, 1948](#)) et la formule (revisitée) de Dale et Chall ([Chall & Dale, 1995](#)) pour l'anglais, et la formule de Henry ([Henry, 1975](#)) pour le français. Ces formules utilisent notamment des variables liées à la fréquence d'occurrence des mots, à la longueur des phrases et des mots ou à la diversité lexicale du texte. Les nouvelles techniques d'apprentissage informatique (*machine learning*) ont permis d'aller beaucoup plus loin dans cette voie en analysant un grand nombre de descripteurs dérivés du texte afin d'obtenir de nouvelles formules. Cependant les nouvelles formules proposées par les linguistes construisant continuent toujours à faire appel à des variables liées aux complexités grammaticales et de vocabulaire ([François, 2009](#)). Tout comme pour la compréhension écrite, les connaissances linguistiques jouent un rôle prépondérant dans la compréhension orale, que ce soit pour la compréhension orale de la langue native ou étrangère ([Buck, 2001](#)) : des connaissances en vocabulaire et grammaire étant nécessaires pour décoder le contenu. Le lien de ces complexités

avec la difficulté de compréhension orale a été montré à plusieurs reprises (Nissan *et al.*, 1995; Carrow-Woolfolk, 1999).

Mais si la complexité intrinsèque du message est importante, sa vocalisation comporte également des aspects importants pour la compréhension orale. Pour l'anglais l'emphase et l'intonation peuvent être déterminantes pour la compréhension du message (Wong & Waring, 2010) en aidant à insister sur les mots importants. Le débit de parole ou l'accent d'un locuteur peuvent également affecter la compréhension (Boyle, 1984). Un débit rapide peut également être une complication pour la compréhension des non natifs (Goh, 1994). Les pauses et les hésitations (aussi appelées disfluences) ont des effets qui varient selon les auditeurs, elles peuvent être informatives pour les auditeurs natifs (Corley & Hartsuiker, 2003) ou porter préjudice à la compréhension des non natifs (Voss, 1979). Chang et Read ont étudié l'impact des accents sur la compréhension orale (Chang & Read, 2008) et ont conclu que la compréhension des auditeurs était affectée par la présence d'accents qui ne leur sont pas familiers.

L'environnement sonore a lui aussi un impact sur la difficulté perçue. Dans l'étude de Boyle (Boyle, 1984), plusieurs professeurs le citent comme un élément important dans la compréhension orale en L1 et L2. Les auditeurs natifs peuvent rencontrer plus de difficultés à comprendre un document audio si les conditions sonores sont mauvaises (Adank *et al.*, 2009). Les effets de l'environnement sonore sont encore plus importants pour des auditeurs non natifs qui sont confrontés à la fois aux conditions d'écoute dégradées et à leurs potentielles lacunes dans la langue cible.

Les études concernant la compréhension de contenus audiovisuels sont moins nombreuses, mais, si l'influence des aspects linguistiques sur la compréhension de contenus audiovisuels n'a visiblement pas été étudiée, on trouve des études s'intéressant à l'influence des sous-titres sur la compréhension de documents audiovisuels. Ce sujet a tout son intérêt dans le cadre de l'apprentissage des langues. Plusieurs tests menés pour déterminer si les sous-titres sont bénéfiques ou nuisibles pour la compréhension, ont principalement conclu que les sous-titres avaient un effet positif en accroissant la compréhension. (Perez *et al.*, 2013; Markham *et al.*, 2001).

Dahl a mené une étude qui a souligné l'importance des gestes dans la facilitation de la compréhension orale L1 et L2 (Dahl & Ludvigsen, 2014) : même si les locuteurs natifs et non natifs ne les exploiteront pas de la même façon, les gestes les aideront à atteindre une meilleure compréhension. En partant de ces observations, l'hypothèse peut être faite que les gestes peuvent faciliter la compréhension des contenus audiovisuels. L'accès à des vidéos qui contiennent des gestes et des indices visuels permet effectivement aux apprenants en langue seconde à mieux réussir la tâche de compréhension orale (Sueyoshi & Hardison, 2005).

Les recherches sur la compréhension de contenu ont mis en avant l'influence des aspects linguistiques sur la compréhension, que ce soit pour la compréhension écrite et orale. Elles ont aussi montré que la qualité de production de la parole (la prosodie, le débit...) et l'environnement sonore ont une influence sur la compréhension de contenu audio. Pour les contenus audiovisuels, la présence d'indices visuels et de sous-titres apparaissent comme des facilitateurs pour la compréhension. Notre étude s'intéresse aux aspects inhérents aux documents audiovisuels qui affectent la compréhension orale : la complexité du vocabulaire, la complexité grammaticale et l'intelligibilité de la parole.

## 3 Matériel et méthodes

### 3.1 Corpus

L'étude porte sur la perception humaine de la difficulté globale d'extraits issus de films. Pour mener cette étude, la création d'un corpus approprié était nécessaire, les sujets de l'expérience devant évaluer la difficulté globale des extraits en fonction des modalités disponibles. Trois modalités ont été exploitées : la modalité audio, la modalité vidéo et la modalité texte. Un corpus composé de 55 extraits de 15 films populaires français a été constitué. Les films ont été choisis pour que le corpus contienne des films de genres et d'époques différents, avec de la variété dans la langue et dans le registre. Les extraits ont été sélectionnés de telle sorte à ce qu'ils répondent à la définition d'une interaction (Traverso, 2013). La compréhension de situations de communication et de leurs éléments contextuels (qui ? quand ? où ? comment ?) étant la première étape menant à la compréhension de contenu audiovisuel. Les scènes d'action, ne comprenant pas de parole ou comportant des interactions trop courtes ont été exclues car elles ont été jugées comme n'étant pas suffisamment pertinentes pour évaluer la difficulté globale. Trois à cinq extraits par films ont été choisis, puis les flux audio et vidéo, ainsi que les transcriptions de ces extraits ont été récupérés. Le corpus compte 55 extraits (2541 secondes), chacun disponible sous trois formats différents :

- texte : il s'agit de la transcription des extraits (7225 mots),
- audio : les pistes sonores seules,
- vidéo : qui inclut les images et le son.

### 3.2 Participants

Les experts de l'étude ont été choisis parmi des professeurs de langue car ils sont souvent confrontés à l'évaluation de documents (de type texte, audio ou vidéo) pour déterminer s'ils sont appropriés pour leurs étudiants en terme de difficulté globale. Quinze professeurs de français langue étrangère (FLE) ont été recrutés, il était requis qu'ils soient francophones natifs avec au moins trois ans d'expérience dans l'enseignement (avec des apprenants de niveaux variés). Ils devaient être familiers avec l'utilisation de documents audiovisuels en classe et être normo-entendants. Les 15 professeurs –13 femmes, 2 hommes, tranche d'âge : 27-63 ans, âge moyen : 37 ans –ont une expérience dans l'enseignement qui varie entre 3 et 40 ans (moyenne : 11 ans, écart-type : 9 ans). Tous les experts ont reçu une compensation financière pour leur participation.

### 3.3 Évaluation

Une interface graphique a été développée puis déployée en ligne, afin de présenter les 55 extraits de films aux différents participants. Les extraits pouvaient être présentés sous cinq conditions différentes, chaque condition correspondant à une combinaison de modalités : texte (T) (seule la transcription de l'extrait était accessible), audio (A), audio+texte (AT), audio+vidéo (AV) et audio+vidéo+texte (AVT). Pour chaque participant, chacun des extraits a été présenté une seule fois sous une des conditions citées ci-dessus. Les participants ont réalisé l'évaluation en ligne, en utilisant leur propre matériel. Il leur a été spécifié de réaliser l'expérience dans un lieu calme, en utilisant un ordinateur et des écouteurs. À la fin de l'évaluation, chacun des 55 extraits de films a été présenté sous chacune des conditions à exactement trois participants.

Les extraits vidéos ont été présentés dans un ordre aléatoire. Pour chacun des extraits, l'interface présentait une zone dédiée à la transcription, un lecteur audio/vidéo ainsi que trois à quatre curseurs, en fonction de la condition présentée. Les participants avaient comme instruction d'utiliser chacun des curseurs mis à leur disposition pour évaluer chacun des extraits en terme de difficulté globale (de 0 –très facile à 100 –très difficile), de complexité du vocabulaire (de 0 –très facile à 100 –très difficile), de complexité grammaticale (de 0 –très facile à 100 –très difficile) et (pour chaque extrait présenté avec la modalité audio) d'intelligibilité de la parole (de 0 –totalement intelligible à 100 –totalement inintelligible). La position initiale de chaque curseur était à 50. L'interface mettait également à disposition des zones de texte où les participants pouvaient laisser un commentaire pour justifier la position donnée à chaque curseur. Les commentaires étaient obligatoires pour justifier l'évaluation de la difficulté globale. Les participants ont été familiarisés avec la tâche d'évaluation lors d'une phase d'entraînement sur cinq extraits de films, qui précédait la véritable évaluation, ils étaient libres de réaliser l'évaluation des extraits en plusieurs fois.

## 4 Résultats

Cette partie traite, dans un premier temps, de l'influence de la complexité du vocabulaire, de la complexité grammaticale et de l'intelligibilité sur la difficulté globale. La dernière partie étudie l'influence des modalités sur la difficulté globale perçue et l'intelligibilité de la parole. La diversité dans la perception des experts pouvant mettre en évidence des facteurs intéressants pour la suite de l'étude, nous n'avons pas procédé à des accords inter-annotateurs.

### 4.1 Relation entre les scores de difficulté globale et les scores de complexité du vocabulaire, de la grammaire et de l'intelligibilité

Dans cette partie les relations entre les scores de difficulté globale et les « sous-scores » (*i.e.* les scores de complexité du vocabulaire, de complexité grammaticale et d'intelligibilité) sont étudiées, à l'aide de corrélations bivariées et de régressions linéaires multiples.

#### 4.1.1 Corrélations bivariées

La table 1 présente les résultats des corrélations par rang de Spearman, calculées entre les scores de difficulté globale et les sous-scores. Dans le cas de l'intelligibilité, les extraits présentés aux participants qui ne contenaient pas la modalité audio ont été ignorés.

TABLE 1 – Corrélations par rang de Spearman entre la difficulté globale et la complexité du vocabulaire, la complexité grammaticale, l'intelligibilité de la parole (\*\*\*)  $p \leq 0,001$

Complexité	Vocabulaire	Grammaire	Intelligibilité parole
Difficulté globale	0,74***	0,56***	0,63***

Des corrélations significatives positives, modérées à fortes, ont été trouvées entre la difficulté globale et les trois sous-dimensions évaluées. Ces résultats montrent que (1) comme cela pouvait être attendu,

plus le vocabulaire et la grammaire sont perçus comme étant complexes plus la difficulté globale perçue augmente et (2) plus la parole est intelligible, plus la difficulté globale perçue diminue. Pour étudier plus en profondeur le lien entre la difficulté globale et les trois sous-dimensions, des régressions linéaires multiples ont été réalisées.

#### 4.1.2 Régressions linéaires multiples

Les scores de difficulté globale et les scores de complexité du vocabulaire, de complexité grammaticale et d'intelligibilité ont été moyennés pour chaque extrait, en prenant en compte les scores donnés sous toutes les conditions. Deux régressions linéaires multiples ont été réalisées en utilisant ces moyennes. La première régression a été calculée en prenant la difficulté globale comme variable dépendante et la complexité du vocabulaire et la complexité grammaticale comme variables indépendantes.

La régression linéaire multiple donne un coefficient de détermination élevé, avec un  $R^2$  ajusté de 0,76. Ceci signifie que le modèle permet d'expliquer 76% de la variance de la difficulté globale. Les coefficients non-standardisés (NsCoef) montrent que la complexité du vocabulaire (NsCoef = 0,69) a plus de poids dans la régression que la complexité grammaticale (NsCoef = 0,34). La figure 1 représente un diagramme de dispersion reliant les scores de difficulté globale moyens prédits aux scores humains de difficulté globale moyens.

Une nouvelle régression linéaire multiple est réalisée pour voir si prendre en considération les scores évaluant l'intelligibilité de la parole permet d'améliorer la prédiction de la difficulté globale. Cette régression prend la difficulté globale comme variable dépendante, et la complexité du vocabulaire, la complexité grammaticale et l'intelligibilité comme variables indépendantes. Pour cette régression, les extraits ayant été notés avec la modalité texte seule ont été ignorés. Rajouter l'intelligibilité comme variable dépendante dans cette seconde régression linéaire permet d'obtenir un meilleur coefficient de détermination, avec un  $R^2$  ajusté de 0,82. La complexité du vocabulaire a toujours plus de poids dans la prédiction de la difficulté globale (NsCoef = 0,55), suivie de la complexité grammaticale (NsCoef = 0,31) et de l'intelligibilité de la parole (NsCoef = 0,28). La relation entre les valeurs prédites et les valeurs réelles sont visibles dans la figure 2.

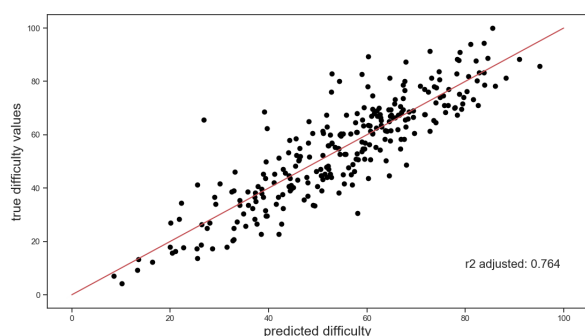


FIGURE 1 – Diagramme de dispersion reliant les évaluations humaines de la difficulté globale aux prédictions de la difficulté globale, en utilisant une régression linéaire multiple avec la complexité du vocabulaire et la complexité grammaticale comme variables indépendantes

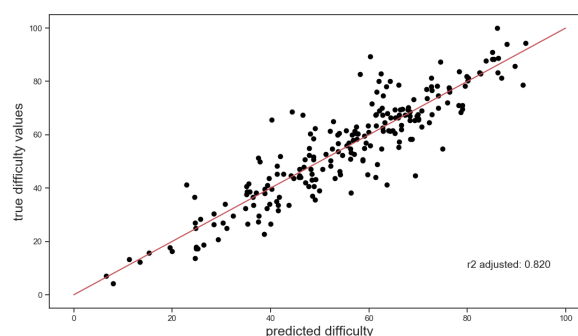


FIGURE 2 – Diagramme de dispersion reliant les évaluations humaines de difficulté globale aux scores prédits de difficulté globale, en utilisant une régression linéaire multiple avec la complexité du vocabulaire, la complexité grammaticale et l'intelligibilité comme variables indépendantes

## 4.2 Influence des modalités

Un autre aspect intéressant à étudier est la variation des scores en fonction des modalités disponibles. Les hypothèses suivantes ont été faites : (1) la difficulté globale perçue sera plus élevée pour la condition A que pour la condition AT et que pour la condition AV, (2) la difficulté globale perçue sera plus élevée pour les conditions AV et AT que pour la condition AVT, (3) l'intelligibilité de la parole augmente si la modalité vidéo et/ou la modalité texte sont combinées avec la modalité audio. Aucune relation n'est attendue entre les modalités et la complexité du vocabulaire et la complexité grammaticale.

La moyenne et l'écart-type des scores de difficulté globale et d'intelligibilité ont été calculés pour chaque condition, pour visualiser l'évolution des scores en fonction des modalités disponibles. Les résultats sont représentés dans les figures 3 et 4.

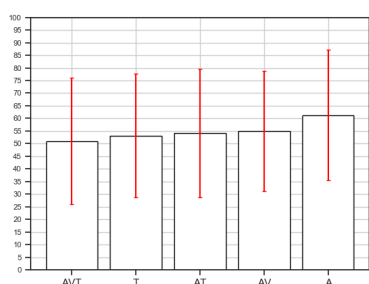


FIGURE 3 – Scores moyens de difficulté globale en fonction des combinaisons de modalités. Les barres d'erreur représentent  $\pm 1$  écart-type

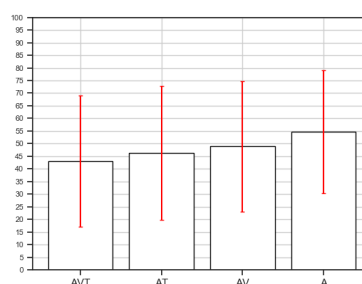


FIGURE 4 – Scores moyens d'intelligibilité de la parole (scores bas = intelligibilité haute) en fonction des combinaisons de modalités. Les barres d'erreur représentent  $\pm 1$  écart-type.

### 4.2.1 Influence des modalités sur la difficulté globale

En premier lieu, il est possible d'observer sur la figure 3 que la moyenne des scores de difficulté globale est la plus élevée dans la condition A, c'est-à-dire si la modalité audio est seule ; dans ce corpus, les extraits présentés avec la modalité audio seule étaient donc les plus compliqués. Comme supposé, les moyennes sont moins élevées quand les modalités vidéo et texte s'ajoutent à la modalité audio seule, la moyenne la plus basse est obtenue pour la condition AVT, c'est-à-dire quand les trois modalités audio, vidéo et texte sont disponibles. Pour ce corpus, avoir accès aux trois modalités en même temps permet de minimiser la difficulté globale perçue.

### 4.2.2 Influence des modalités sur l'intelligibilité de la parole

Avoir accès à la vidéo et au texte devrait représenter un avantage pour la compréhension : si l'accès à la vidéo ne permet pas totalement de désambiguïser ce qui est entendu, la présence du texte permet de définitivement éliminer les problèmes d'inintelligibilité (même s'il est possible qu'un surplus d'information amène une surcharge cognitive). Cette théorie peut être supportée par les observations faites sur l'évolution de la moyenne des scores d'intelligibilité en fonction des modalités disponibles. La figure 4 montre que les extraits vidéos présentés avec la modalité audio seule, sont



perçus comme moins intelligibles par les participants et que l'ajout des modalités vidéo et/ou texte améliore l'intelligibilité. Les extraits vidéos les plus intelligibles sont ceux qui ont été présentés avec les trois modalités audio, vidéo et texte. Ces résultats montrent que, pour ce corpus, combiner les modalités audio, vidéo et texte permet d'optimiser l'intelligibilité.

## 5 Discussion et conclusions

Dans le cadre d'une étude de la compréhension de contenu audiovisuel, un corpus composé d'évaluations subjectives de 55 extraits extraits de 15 films, réalisées par 15 professeurs de français langue étrangère, a été constitué. Ces évaluations ont permis d'étudier les facteurs qui influent sur la difficulté perçue mais aussi d'étudier l'influence des modalités sur la compréhension.

Concernant l'influence de la complexité du vocabulaire, de la complexité grammaticale et de l'intelligibilité de la parole sur la difficulté globale, au niveau du corpus, il a été confirmé qu'une corrélation positive existait entre ces facteurs et la difficulté globale. Des régressions linéaires multiples, prenant la difficulté globale comme variable dépendante et la complexité du vocabulaire, la complexité grammaticale et l'intelligibilité comme variables indépendantes, confirment que la combinaison de ces trois variables permet de construire un modèle prédisant efficacement la difficulté globale, avec un  $R^2$  ajusté atteignant une valeur de 0,82. Le fait que le  $R^2$  ajusté n'atteigne pas la valeur de 1 peut s'expliquer par le fait que certains éléments jouant sur la difficulté globale n'ont pas été pris en considération. Des facteurs secondaires affectant la difficulté globale (par exemple des facteurs liés à l'aspect cognitif) n'ont pas été inclus dans cette étude. De plus, comme les évaluations ont été réalisées par des humains, d'autres facteurs, comme la fatigue des participants, ont pu influencer les évaluations.

En ce qui concerne l'influence des modalités sur la difficulté globale, il a été montré que les extraits présentés avec la modalité audio seule ont été perçus comme étant les plus difficiles par les participants. Rajouter les combinaisons vidéo et texte diminue la difficulté globale perçue et améliore l'intelligibilité de la parole : les extraits les plus faciles et les plus intelligibles du corpus étant ceux qui étaient présentés sous la condition AVT. La présence du texte doit permettre de contourner les problèmes liés à l'intelligibilité. La présence de la vidéo peut aider à améliorer l'intelligibilité de part les indices que peuvent donner les gestes, les expressions faciales et les mouvements des lèvres. En résumé, l'exploitation de toutes les modalités semble permettre de maximiser l'intelligibilité et de minimiser la difficulté globale perçue.

Ces conclusions résultent d'une étude subjective qui pourra conduire au développement d'un outil de prédiction automatique de la difficulté globale et à l'obtention d'une mesure objective de cette difficulté. La prochaine étape consistera à comparer les scores recueillis avec les scores obtenus automatiquement à partir de paramètres extraits des modalités audio, texte et vidéo, et à différentes modélisations pour voir si la complexité du vocabulaire, de la grammaire et l'intelligibilité peuvent être prédites. L'intérêt de cette étude était de recueillir des scores reflétant la perception d'experts de la difficulté des documents audiovisuels et de déterminer les facteurs influençant la compréhensibilité tout en incluant les variations de perception de chacun des experts. Les critères mis en évidence pourront être intégrés dans les travaux futurs sur la prédiction de la difficulté globale, à leur tour évalués par des enseignants de langue ou des apprenants.

## Références

- ADANK P., EVANS B., STUART-SMITH J. & SCOTT S. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology : Human perception and performance*, **35**(2), 520.
- BOYLE J. (1984). Factors affecting listening comprehension. *ELT Journal*, **38**(1), 34–38.
- BUCK G., Éd. (2001). *Assessing Listening*. Cambridge University Press.
- CARROW-WOOLFOLK E. (1999). Comprehensive assessment of spoken language. *Bloomington, MN : Pearson Assessment*.
- CHALL J. S. & DALE E., Éd. (1995). *Readability revisited : The new Dale-Chall readability formula*. Brookline Books.
- CHANG A.-S. & READ J. (2008). Reducing listening text anxiety through various forms of listening support. *TESL-EJ*, **12**(1), 1–25.
- CORLEY M. & HARTSUIKER R. (2003). Hesitation in speech can... um... help a listener understand. *Proceedings of the Annual Meeting of the Cognitive Science Society*, **25**(25), 276–281.
- DAHL T. & LUDVIGSEN S. (2014). How I see what you're saying : The role of gestures in native and foreign language listening comprehension. *The Modern Language Journal*, **98**(3), 813–833.
- FLESCH R. (1948). A new readability yardstick. *Journal of Applied Psychology*, **32**(3), 221–233.
- FRANÇOIS T. (2009). Modèles statistiques pour l'estimation automatique de la difficulté de textes de FLE. *Actes de RECITAL 2009*.
- GOH C. (1994). How much do learners know about the factors that influence their listening comprehension? *Hong Kong Journal of Applied Linguistics*, **4**(1), 17–42.
- HENRY G., Éd. (1975). *Comment mesurer la lisibilité*. Labor.
- LIVELY B. & PRESSEY S. (1923). A method for measuring the 'vocabulary burden' of textbooks. *Educational Administration and Supervision*, **9**, 938–398.
- MARKHAM P., PETER L. & MCCARTHY T. (2001). The effects of native language vs target language captions on foreign language students' dvd video comprehension. *Foreign Language Annals*, **34**(5), 439–445.
- NISSAN S., DEVINCENZI F. & TANG K. (1995). An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension. *ETS Research Report Series*, **2**, i–42.
- PEREZ M., NOORTGATE W. & DESMET P. (2013). Captioned video for L2 listening and vocabulary learning : A meta-analysis. *System*, **41**(3), 720–739.
- SUEYOSHI A. & HARDISON D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, **55**, 661–699.
- TRAVERSO V., Éd. (2013). *L'analyse des conversations*. Armand Colin.
- VOSS B. (1979). Hesitation phenomena as sources of perceptual errors for non-native speakers. *Language and Speech*, **22**(2), 129–144.
- WONG J. & WARING H., Éd. (2010). *Conversation Analysis and Second Language Pedagogy*. Taylor & Francis.