

Problématique

Nous avons observé un changement de paradigme dans le domaine de l'évaluation de la prononciation en L2 : ce n'est plus la distance avec une prononciation native qui importe, mais l'efficacité de la communication en termes de compréhension. Ce qui compte, c'est que l'auditeur comprenne le locuteur et qu'il le comprenne avec autant de facilité que possible. On parle d'« intelligibilité » pour qualifier le niveau de compréhension de cet auditeur vis-à-vis de la production du locuteur, et de « compréhensibilité » pour qualifier le degré d'aisance avec lequel il parvient à comprendre.

Les descripteurs de compétences linguistiques internationaux comme le Cadre Européen Commun de Référence, ainsi que les grilles d'évaluation de production orale des principaux tests certificatifs d'anglais, mettent en effet clairement en avant l'importance de la compréhension. Le niveau de l'apprenant est maintenant déterminé en fonction de son intelligibilité du point de vue de l'auditeur. Un seuil particulièrement important est décrit au niveau B2 ou équivalent, où le locuteur devient globalement fluide et facilement compréhensible.

Cependant, les descripteurs manquent de définitions explicites de ce qui constitue une parole fluide et compréhensible, et le jugement reste à la charge de l'interprétation de l'évaluateur. On constate pourtant que leur évaluation est généralement consistante parmi les évaluateurs, qu'ils soient experts ou non, et qu'ils participent ou non directement à l'interaction.

En parallèle, les outils d'évaluation automatique de la prononciation reposent soit sur des paramètres de bas niveau corrélés avec le niveau global de l'apprenant mais toutefois généralement peu pertinents sur le plan pédagogique – comme le débit de parole ou la fréquence des pauses –, soit sur la mesure d'une distance vis-à-vis d'un modèle natif, sans chercher à identifier les éléments qui réduisent la compréhensibilité. De plus, il est important de pouvoir évaluer la performance de l'apprenant en contexte spontané, en interaction, afin d'estimer sa compétence en situation de communication.

Il apparaît donc souhaitable de développer des systèmes automatiques capables d'évaluer la production orale spontanée en se basant sur des phénomènes linguistiques spécifiquement susceptibles d'entraver la compréhension.

Facteurs linguistiques clés

De nombreux facteurs peuvent impacter l'intelligibilité et la compréhensibilité, depuis la précision phonologique de l'énoncé du locuteur, à l'implication de l'auditeur dans la conversation, en passant par sa familiarité avec le sujet abordé. Parmi ces facteurs, plusieurs dépendent avant tout de la performance linguistique du locuteur, bien que leur effet puisse varier selon la compétence de compréhension de l'auditeur. Nous proposons de nous concentrer ici sur les deux éléments suivants :

- **Distribution des pauses** : Les pauses situées au niveau de frontières syntaxiques de haut niveau, typiquement entre propositions, tendent à améliorer la compréhension et reflètent une meilleure planification de l'énoncé. Par exemple, une pause après une conjonction de coordination ou avant une relative permet de structurer l'information de manière plus accessible pour l'auditeur. En revanche, celles situées à l'intérieur d'unités syntaxiques de petite taille sont souvent perçues comme des disfluences qui rendent la compréhension difficile.
- **Rythme et accentuation lexicale** : Le rythme de la parole joue un rôle important dans la compréhensibilité du locuteur. En anglais, il est défini en grande partie par l'accentuation des voyelles, qui, au-delà de jouer un rôle lexical permettant une meilleure reconnaissance des mots, permet aussi une segmentation de la parole à un plus bas niveau : l'accent se trouve fréquemment en tête de mot et en début de groupe prosodique, et les mots grammaticaux ne sont généralement pas accentués. Par ailleurs, les voyelles non accentuées sont souvent réduites, créant un contraste prosodique marqué avec les voyelles accentuées.

Les récentes avancées en traitement automatique de la parole ouvrent de nouvelles perspectives pour l'évaluation de la production orale spontanée des apprenants de langues étrangères. En combinant reconnaissance automatique de la parole et analyses syntaxiques, il devrait être possible de détecter les pauses et d'identifier leur position par rapport aux frontières syntaxiques. Cette approche permettrait de considérer les pauses relativement à la structure syntaxique de l'énoncé et ainsi de valoriser celles qui sont placées à des frontières stratégiques, ou de ne pénaliser que celles qui ne le sont pas.

De même, en exploitant des techniques de détection des noyaux syllabiques et d'analyses acoustiques, il devrait être possible d'identifier les syllabes acoustiquement proéminentes susceptibles d'être perçues comme accentuées par les auditeurs, de comparer leur position avec la position attendue de l'accent lexical selon un dictionnaire de référence, mais aussi de quantifier le contraste acoustique entre les syllabes accentuées et les syllabes réduites.

Questions de recherche

L'objectif de cette thèse est de répondre à trois questions principales :

1. Peut-on concevoir un outil d'annotation automatique de la parole spontanée L2 capable de caractériser la distribution des pauses et les patterns accentuels qui rendent la compréhension difficile ?
2. Observe-t-on des différences significatives entre les locuteurs de niveaux B1 et B2 en termes de distribution syntaxique des pauses et d'accentuation lexicale ?
3. Peut-on mesurer dynamiquement l'impact de ces facteurs sur la perception de compréhension chez les auditeurs ?

Par la réalisation d'un outil de ce type, notre objectif est de proposer un complément d'informations quantifiables pour les évaluateurs, mais aussi un outil de diagnostic pour les enseignants et les apprenants, en exploitant le potentiel des outils actuels de traitement automatique de la parole.

Hypothèses

En appliquant des outils de traitement automatique à un corpus de productions orales spontanées d'apprenants francophones et japonophones de niveaux B1 et B2, plusieurs comparaisons peuvent être envisagées :

- **Distribution des pauses** : Les locuteurs B2, perçus comme plus fluides et compréhensibles, devraient produire une proportion plus élevée de pauses en frontières syntaxiques de haut niveau (ex. : entre propositions), et moins de pauses en frontières de bas niveau (ex. : au sein d'unités syntaxiques réduites).
- **Accentuation lexicale** : Les productions devraient refléter une influence variable des langues sources. Les francophones, issus d'une langue à accent fixe marqué par un allongement de la syllabe finale, devraient présenter une tendance à rallonger systématiquement les syllabes en fin de mot et faire peu varier la f_0 et l'intensité. Pour les locuteurs japonophones, en revanche, nous nous attendons à observer moins de difficultés à positionner correctement l'accent, puisque le japonais possède également un accent lexical, mais toutefois un contraste plus marqué par une variation de la f_0 , que par celle de l'intensité ou de la durée.

Enfin, nous nous questionnons sur l'effet concret que peuvent avoir les pauses situées en frontières syntaxiques de bas niveau, et les patterns accentuels non adéquats, sur la perception de difficulté de compréhension chez l'auditeur. Nous faisons l'hypothèse que ces phénomènes ont un impact direct et observable sur la perception de compréhensibilité, et qu'ils peuvent être mesurés à l'aide d'un protocole d'évaluation dynamique semblable à celui utilisé par [Nagle et al. \(2019\)](#). Le jugement de difficulté perçue par les auditeurs devrait avoir tendance à augmenter à la suite de ces phénomènes, et au contraire à diminuer, ou au moins ne pas augmenter, à la suite de pauses en frontières syntaxiques de haut niveau et de patterns accentuels adéquats.