

Conclusion générale

La production et la compréhension de l'oral sont deux compétences étroitement liées. Dans une situation de communication, produire de la parole vise nécessairement à être compris. Évaluer la compétence de production orale à travers la capacité du locuteur à se faire comprendre prend alors tout son sens. Dès lors, il semble plus pertinent de considérer cette compétence comme un tout, sans isoler l'aspect spécifique de la prononciation, qui n'est finalement qu'une forme donnée à la parole et ne peut être évaluée de manière intrinsèque.

Si la compréhensibilité du locuteur dépend pour une grande part de la capacité de compréhension de l'auditeur, elle peut être facilitée par certains aspects spécifiques de la production orale. Par exemple, une segmentation stratégique du flux de parole peut aider l'auditeur à structurer et à traiter l'information reçue, mais également lui permettre de réagir, de signaler au locuteur qu'il écoute ou qu'il comprend. La communication est ainsi une co-construction permanente entre le locuteur et l'auditeur, et évaluer les compétences de l'apprenant devrait porter sur sa capacité à co-construire cette communication, tant sur les aspects de compréhension (comprendre) que de compréhensibilité (se faire comprendre).

Mais alors, est-ce qu'évaluer la production orale de manière automatique est une entreprise viable ? Comment évaluer la production d'un apprenant autrement que par le biais de la compréhension de son interlocuteur ?

Il y a trois ans, au début de cette thèse, l'idée de proposer un environnement de conversation où l'apprenant pourrait dialoguer avec un avatar numérique capable d'interagir avec lui de manière similaire à un humain semblait encore appartenir à un futur lointain. Si ce n'est pas encore tout à fait une réalité, l'émergence récente des grands modèles de langage laisse entrevoir des perspectives réalistes d'environnements numériques pour la pratique et l'évaluation de la production orale spontanée en L2. Des entreprises comme LangX⁷ proposent déjà des systèmes conversationnels spécialisés dans l'évaluation des compétences en langues. Toutefois, ces technologies ne

⁷<https://speaking.langx.ai/>

permettent que de proposer un environnement de production de parole : les questions de quoi et comment évaluer restent entières.

L'évaluation peut porter sur des aspects propres à la production du locuteur, mais également sur la dynamique des échanges, sur la capacité du locuteur à se faire comprendre et à s'adapter en fonction des réactions de son interlocuteur. Elle peut se faire dans différentes situations de communication, plus ou moins formelles, avec des interlocuteurs plus ou moins familiers avec la parole L2, ou encore avec le sujet abordé.

Concernant plus spécifiquement les aspects de production, il apparaît que la segmentation du flux de parole (à travers la distribution des pauses) et le rythme général (via les phénomènes d'accentuation) ont un impact direct sur la compréhension. Toutefois, ces deux paramètres doivent être appréhendés comme des variables continues et relatives. Il s'est avéré en effet plus pertinent de considérer la position des pauses en fonction du degré d'importance de la frontière syntaxique où elles se trouvent, plutôt que vis-à-vis du type de constituants grammaticaux qui s'y terminent ou commencent. De même, la mesure du contraste accentuel entre les syllabes s'est révélée plus informative que la simple identification des syllabes proéminentes. En outre, une pause de bas niveau syntaxique ou une accentuation non prescriptive ne constituent pas en elles-mêmes des obstacles à la compréhension ; c'est leur récurrence ou leur accumulation, combinée à d'autres facteurs, qui peut accroître l'effort de compréhension requis par l'auditeur. Ce phénomène illustre bien le « cocktail de l'intelligibilité » décrit par Zielinski (2006). La distribution des pauses et l'accentuation font partie d'un ensemble de paramètres qu'il convient de prendre en compte dans l'évaluation, mais celle-ci doit se faire à un niveau global, en combinant les aspects lexicaux, grammaticaux ou encore discursifs, mais également en tenant compte du contexte de la situation de communication et des réactions de l'auditeur.

Perspectives pour SELF

Quelles sont les perspectives pour le dispositif d'évaluation SELF ? Deux objectifs principaux se dessinent. Le premier est d'intégrer la production orale dans le test de positionnement actuellement déployé. Cela passera par l'ajout de courtes tâches d'élicitation de parole, à partir desquelles plusieurs mesures seront effectuées. Ces mesures combineront des aspects de fluence (par exemple, le score de distribution syntaxique des pauses DSP_n proposé dans cette thèse), de rythme (notamment via la mesure du contraste prosodique moyen \bar{C}), ainsi que des aspects lexicaux et syntaxiques (précision, diversité) ou pragmatiques (adéquation au contexte). Ces indicateurs contribueront à estimer un niveau de compétence qui viendra enrichir les scores actuellement calculés.

Cependant, si l'objectif immédiat est d'estimer le niveau des apprenants afin de les orienter vers des groupes adaptés, l'ambition à plus long terme est de développer un module d'évaluation formative de la production orale. Ce module devra permettre d'établir un diagnostic sur la base de ces différentes mesures, dans une situation de communication plus réaliste et sur une durée d'évaluation plus longue. L'objectif principal sera d'offrir à l'apprenant une estimation de son niveau de compréhensibilité dans un contexte donné et d'identifier les paramètres prioritaires à travailler pour améliorer ses performances en communication.

Si ces avancées technologiques permettent d'imaginer une évaluation entièrement automatisée, il convient de souligner que celle-ci ne pourrait porter que sur des aspects superficiels de la communication. Les mesures évoquées ici constituent un ensemble d'indices susceptibles de favoriser la réussite de la communication ou de refléter une communication réussie, permettant ainsi d'estimer la compétence de production orale du locuteur. En revanche, une évaluation approfondie ne peut être envisagée que grâce à l'expertise d'un évaluateur humain, capable, contrairement à la machine, d'interpréter la production en profondeur – tant sur le plan du sens que par une combinaison plus nuancée des différents paramètres.

Une évaluation automatisée, limitée à des paramètres de surface, peut néanmoins représenter un complément précieux, en allégeant la charge de l'évaluateur humain et en lui permettant de se concentrer sur l'interprétation globale et l'établissement du jugement final. Cela explique pourquoi une automatisation totale demeure inadéquate dans le cadre d'une évaluation certificative, où les enjeux sont souvent importants. En revanche, dans un contexte de formation, le potentiel de l'évaluation automatique est considérable. Il nous revient alors de nous approprier ces technologies et de les adapter à nos besoins afin d'en tirer pleinement parti.

Nous concluons par les mots d'un chercheur qui a profondément influencé notre réflexion tout au long de ce parcours doctoral : *“Rather than a false fight over an already too small piece of the language teaching pie, [technology] is a way to expand the pie so that more teachers and learners can enjoy their own use of spoken language”* (Levis, 2007, p. 197).