

Peut-on évaluer la compréhensibilité de la parole sans référence quant aux intentions de communication du locuteur ? Une étude auprès d'apprenants germanophones de FLE

Verdiana De Fino^{1,2} Isabelle Ferrané¹ Julien Pinquier¹ Lionel Fontan²

(1) IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France

(2) Archean LABS, Montauban, France

verdiana.defino@irit.fr, isabelle.ferrane@irit.fr, lfontan@archean.tech,
julien.pinquier@irit.fr

RÉSUMÉ

En didactique des langues étrangères, la compréhensibilité des énoncés produits par les apprenants est le plus souvent évaluée de manière subjective, à l'aide d'échelles qualitatives. Très souvent, ces évaluations sont menées sans que l'évaluateur ne soit informé du contenu sémantique du message que souhaitait transmettre l'apprenant. L'évaluateur peut donc ignorer des divergences entre ce dernier et sa propre interprétation de l'énoncé – avec pour conséquence une surestimation de la compréhensibilité. Dans cette étude, nous vérifions l'existence d'un tel biais en demandant à 80 francophones natifs d'évaluer la compréhensibilité d'énoncés produits par neuf apprenants germanophones de français lors d'une tâche de traduction. L'évaluation est conduite sans référence (condition « *a priori* »), et en prenant connaissance d'une traduction de référence (condition « *a posteriori* »). Les résultats démontrent que les scores de compréhensibilité sont significativement plus élevés dans la condition *a priori* que dans la condition *a posteriori*, avec une taille d'effet moyenne.

ABSTRACT

Can speech comprehensibility be assessed without any reference as to the communicative intent of the speaker? A study in German speakers of French as a foreign language

The comprehensibility of speech utterances produced by non-native speakers of foreign languages is generally assessed using subjective ratings on qualitative scales. Often, raters are not informed about the semantic content of the messages that the speakers intended to transmit. The raters can thus overestimate comprehensibility, given that they cannot be aware of discrepancies between the communicative intents of the speakers and their own understanding of the utterances. The present study aimed at verifying the existence of this bias by asking 80 native-French speakers to assess the comprehensibility of utterances produced by nine German learners of French during a translation task. The assessment took place without providing to the rater any reference as to the communicative intent of the speaker (*a priori* assessment) and after providing a reference translation (*a posteriori* assessment). As was hypothesized, the results show that *a priori* ratings are significantly higher than *a posteriori* ratings, with a medium effect size.

MOTS-CLÉS : Compréhensibilité de la parole, évaluation, biais, L2.

KEYWORDS: Speech comprehensibility, assessment, bias, L2.

1 Introduction

En didactique des langues étrangères (L2), il a pendant longtemps été considéré que l'objectif ultime de l'enseignement était, pour l'apprenant, d'atteindre le niveau de compétence d'un locuteur natif, et ce y compris au niveau phonético-phonologique – avec la production d'une parole parfaitement intelligible, sans qu'aucun accent étranger ne soit perceptible (Derwing, 2003). Cependant, il s'agit là d'un objectif que très peu d'apprenants parviennent à réaliser, en particulier lorsque le processus d'acquisition débute après l'enfance (Flege, 1988). Peu à peu, cette perspective a donc évolué et la majorité des didacticiens considèrent aujourd'hui qu'il est bien plus important pour l'apprenant de parvenir à une bonne *compréhensibilité* lors d'interactions avec les locuteurs de la communauté cible que de parvenir à s'exprimer « sans accent » et avec un niveau de performance égal à celui d'un locuteur natif (Derwing & Munro, 2009; Munro & Derwing, 2011).

Par conséquent, dans le domaine de l'acquisition/apprentissage des langues, la performance des locuteurs est souvent évaluée à travers la compréhensibilité des énoncés qu'ils produisent dans la langue cible. Pour cela, la méthode la plus utilisée (le « *gold standard* » actuel) consiste à demander à des auditeurs natifs de la L2 d'évaluer le degré de difficulté qu'ils éprouvent à comprendre lesdits énoncés (Derwing & Munro, 2005). De nombreuses études ont ainsi recours à des échelles perceptives d'évaluation de la compréhensibilité, appliquées à l'évaluation d'énoncés produits dans des tâches de production orale plus ou moins contraintes : de la lecture oralisée (Ludwig & Mora, 2017) à la parole recueillie durant des tâches de production plus spontanées comme des tâches d'argumentation (Suzuki & Kormos, 2020).

Demander à un auditeur d'estimer, sur une échelle, le degré de difficulté qu'il a éprouvé pour comprendre un message suppose que celui-ci connaisse les intentions de communication du locuteur, autrement dit, qu'il connaisse le contenu sémantique, discursif, voire pragmatique, que le locuteur souhaitait véhiculer avec son message (Fontan, 2012; Fontan *et al.*, 2013). Or, dans une grande partie des travaux en L2, les auditeurs évaluent la compréhensibilité sans aucune connaissance quant au contenu du message que le locuteur souhaitait transmettre. C'est en particulier le cas pour les études reposant sur des énoncés produits dans des tâches de production orale (semi-)spontanées, pour lesquelles il n'existe pas de « script » correspondant aux énoncés que l'apprenant devait réaliser. Cependant, cette méthode est aussi appliquée pour l'évaluation de la parole lue. Pour illustration, la table 1 recense des études relativement récentes dans lesquelles la compréhensibilité d'apprenants de L2 a été évaluée sur différentes échelles perceptives, sans qu'aucune information n'ait été donnée aux évaluateurs quant aux intentions de communication des apprenants.

Le fait que l'auditeur ne dispose d'aucune information quant au message que l'apprenant souhaitait transmettre peut représenter un biais pour l'évaluation de la compréhensibilité de la parole. En effet, les « erreurs » de prononciation, de choix lexicaux, ou de constructions morphosyntaxiques commises par les apprenants peuvent conduire à la production d'énoncés pouvant être perçus par les évaluateurs comme étant tout à fait corrects. Dans ce cas, ces énoncés peuvent recevoir des scores de compréhensibilité élevés, même s'ils ne correspondent pas aux messages que les apprenants souhaitaient transmettre. Pour prendre le seul exemple d'une erreur de prononciation, il est possible qu'un apprenant produise une phrase perçue comme « Je veux deux cafés », alors qu'il souhaitait dire « Je veux du café », mais que la réalisation du déterminant « du » (/dy/) soit perçue comme « deux » ([dø]) à cause de son système interphonologique. Dans ce cas, si l'évaluateur ne connaît pas le message que souhaitait transmettre l'apprenant, il pourra alors sous-estimer sa difficulté à comprendre l'énoncé.

Étude	Tâche de production orale	Outil d'évaluation de la compréhensibilité
Bergeron & Trofimovich (2017)	Parole spontanée (narration d'histoire imagée et entretien semi-dirigé)	Échelle continue
Crowther <i>et al.</i> (2015)	Parole spontanée (narration d'histoire imagée)	Échelle continue
Hansen Edwards <i>et al.</i> (2018)	Lecture de texte Parole spontanée (entretien semi-dirigé)	Échelle à 9 points
Kang (2010)	Parole spontanée (entretien semi-dirigé)	Échelle à 7 points
Kennedy & Trofimovich (2008)	Lecture oralisée de phrases	Échelle à 9 points
Ludwig & Mora (2017)	Lecture oralisée de phrases	Échelle à 7 points
Nagle & Huensch (2020)	Parole spontanée (entretien semi-dirigé)	Échelle continue
Saito <i>et al.</i> (2017)	Parole spontanée (narration d'histoire imagée)	Échelle continue
Saito <i>et al.</i> (2023)	Parole spontanée (description d'image)	Échelle à 9 points
Suzukida & Saito (2021)	Parole spontanée (narration d'histoire imagée)	Échelle à 9 points

TABLE 1 – Exemples d'études dans lesquelles la compréhensibilité d'apprenants de L2 a évaluée sans informer les évaluateurs du contenu sémantique des messages que les apprenants avaient pour intention de transmettre.

La présente étude a pour objet de vérifier l'existence d'un tel biais lors de l'évaluation de la compréhensibilité de la parole d'énoncés produits par des apprenants germanophones de français langue étrangère (FLE), recueillis pendant une tâche de traduction orale de phrases cibles. À cette fin, un protocole permettant d'évaluer la compréhensibilité des énoncés sans aucune connaissance du message cible, et avec la mise à disposition d'une traduction de référence en français a été mis en place. Nos hypothèses sont celles (i) d'une surestimation de la compréhensibilité des énoncés lorsqu'ils sont présentés sans aucune référence, et (ii) de meilleurs accords inter-évaluateurs après la mise à disposition des traductions de référence en comparaison des accords observés pour les évaluations réalisées *a priori*.

2 Méthode

2.1 Stimuli de parole

Neuf locuteurs germanophones ont participé à l'étude. Il s'agit de neuf apprenants de FLE (2 femmes, 7 hommes), dont les niveaux de compétences selon le Cadre Européen Commun de Référence pour

les Langues (CECRL, [Conseil De l'Europe, 2001](#)) s'échelonnaient de A1 à B1.

Les apprenants ont été enregistrés lors de la traduction orale de 40 phrases cibles présentées en allemand. Ces 40 phrases, élaborées avec l'aide de deux enseignants allemands de FLE, correspondaient à des phrases dans un registre de langue courant, et étaient de nature déclarative. Chaque phrase était destinée à susciter des erreurs lexicales, syntaxiques ou morphosyntaxiques fréquentes chez les apprenants germanophones de FLE. Ainsi, une erreur courante au niveau lexical est d'utiliser le verbe « recevoir » au lieu d'« obtenir » (*p. ex.* « J'ai reçu mon diplôme » au lieu de « J'ai obtenu mon diplôme »), au niveau syntaxique d'utiliser un syntagme nominal en lieu et place d'un syntagme prépositionnel comme complément au verbe « répondre » (*p. ex.* « Nous devons répondre cette question » au lieu de « Nous devons répondre à cette question ») et au niveau morphosyntaxique d'utiliser des formes comme « vieux » ou « beau » devant un nom masculin commençant par une voyelle (*p. ex.* « C'est un vieux homme » au lieu de « C'est un vieil homme »).

Les énoncés produits par les apprenants ont été enregistrés dans une salle calme de l'Université Ostfalia à Wolfenbüttel (Allemagne), à l'aide d'un microphone casque Jabra (Copenhague, Danemark) Evolve 20 HSC016. Le niveau sonore des 360 fichiers audio produits a été normalisé avant les évaluations.

2.2 Évaluation de la compréhensibilité de la parole

Quatre-vingts évaluateurs (26 femmes, 54 hommes) ont été recrutés pour évaluer le degré de compréhensibilité des énoncés produits par les apprenants. Les critères d'inclusion dans l'étude étaient les suivants : francophone natif, sans trouble auditif connu, et âgé entre 18 et 40 ans (pour limiter la probabilité de troubles auditifs liés à l'âge dont les évaluateurs n'auraient peut-être pas conscience, [Cruickshanks et al., 1998](#)).

Les évaluations ont pris place de manière individuelle dans des pièces calmes. Les stimuli étaient diffusés dans un casque Jabra (Copenhague, Danemark) Evolve 20 HSC016, connecté à la carte son d'un ordinateur portable MacBook Pro 13 pouces. Sur l'écran d'ordinateur, une interface graphique développée en langage de programmation Python avec la librairie Streamlit 3¹ était présentée (voir Figure 1).

Avant de démarrer la procédure, chaque évaluateur était familiarisé avec le concept de compréhensibilité de la parole. En accord avec la définition proposée par V. Woisard et collègues (« Capacité de l'auditeur à interpréter le sens du message oral produit par un locuteur, sans tenir compte de la précision ou de la justesse phonétique ou lexicale » ; [Woisard et al., 2013](#)), nous avons en particulier insisté sur le fait que l'évaluation portait sur la capacité à *interpréter le sens* des énoncés, et qu'il ne fallait pas, pour cela, tenir compte d'« erreurs » phonétiques ou linguistiques (par exemple, une prononciation déficiente) qui n'entraveraient pas cet accès au sens. Chaque évaluateur était ensuite familiarisé avec l'échelle à 5 points utilisée pour l'évaluation de la compréhensibilité, échelle qui s'étendait de 1 (« compréhensibilité nulle ») jusqu'à 5 (« compréhensibilité totale »).

Scores de compréhensibilité *a priori*

Pour chaque énoncé, chaque participant devait d'abord évaluer la compréhensibilité de la parole sans aucune indication quant au contenu sémantique cible du message produit par l'apprenant (voir cadre 1

1. <https://streamlit.io/>

Évaluation de la compréhensibilité de la parole

Écouter

Écrivez ce que vous entendez en utilisant de vrais mots (mots du dictionnaire)

j'ai donné rendez vous à un livreur

Évaluation de la compréhensibilité

Modérée

Nulle Totale

Valider

La traduction attendue était :

Je veux vous montrer un livre

Maintenant que vous avez lu la phrase cible, évaluez de nouveau la compréhensibilité du message entendu précédemment

Nulle

Nulle Totale

Valider

Cadre 1 :
Écoute du stimulus

Cadre 2 :
Évaluation de la compréhensibilité *a priori*

Cadre 3 :
Évaluation de la compréhensibilité *a posteriori*

FIGURE 1 – Interface logicielle utilisée pour l'évaluation de la compréhensibilité de chaque enregistrement. L'exemple correspond à de vraies évaluations obtenues pour une traduction orale de la phrase « Ich möchte Ihnen ein Buch zeigen » (« Je veux vous montrer un livre »), perçue par l'évaluateur comme « j'ai donné rendez-vous à un livreur ».

de l'interface). Après l'écoute de la production orale, l'évaluateur avait pour consigne de retranscrire la phrase entendue², et d'attribuer un score de compréhensibilité sur l'échelle à 5 points (voir cadre 2 de l'interface). Nous nommons ces évaluations « Scores de compréhensibilité *a priori* ».

Scores de compréhensibilité *a posteriori*

Une traduction de référence de l'énoncé en français, établie par un consensus entre les enseignants qui ont défini le jeu de phrases à traduire, était ensuite présentée à l'évaluateur. Après la prise de connaissance de cette traduction de référence, l'évaluateur devait à nouveau attribuer un score de compréhensibilité sur l'échelle à 5 points (voir cadre 3 de l'interface). Nous nommons ces secondes évaluations « Scores de compréhensibilité *a posteriori* ».

Les évaluateurs ne pouvaient écouter qu'une seule fois les enregistrements, avant l'évaluation *a priori*. Aucune écoute supplémentaire entre les évaluations *a priori* et *a posteriori* n'était possible. De plus, afin d'éviter les effets d'entraînement (habitude au locuteur), chaque évaluateur a évalué neuf productions orales (une par apprenant). Cela a donné lieu à 18 scores de compréhensibilité : neuf *a priori*, neuf *a posteriori*. Enfin, l'attribution des fichiers audios aux différents évaluateurs a été réalisée de manière à ce que chaque couple d'évaluateurs ait à évaluer un même ensemble de neuf fichiers audio, permettant ainsi le calcul d'accords inter-évaluateurs.

Au total, 1440 scores de compréhensibilité ont ainsi été recueillis, correspondant aux 2 conditions

2. Donnée que nous n'exploitons pas pour la présente étude.

d'évaluation (*a priori* vs. *a posteriori*) $\times$ 9 fichiers audio $\times$ 80 évaluateurs.

3 Résultats

3.1 Scores de compréhensibilité *a priori* et *a posteriori*

La Figure 2 présente les distributions des scores de compréhensibilité *a priori* (à gauche) et *a posteriori* (à droite).

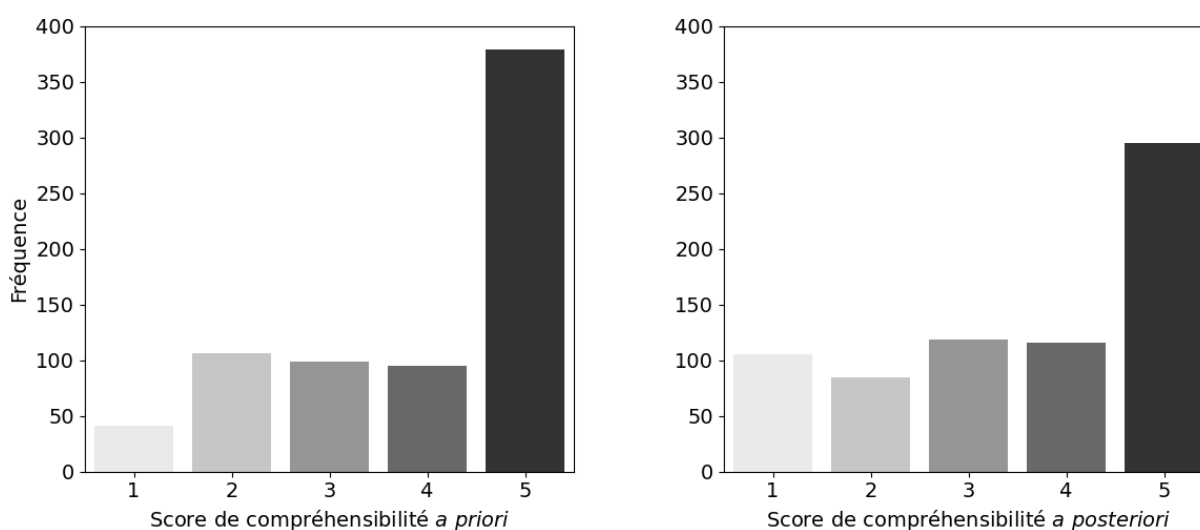


FIGURE 2 – Distribution des scores de compréhensibilité attribués *a priori* et *a posteriori*.

Comme nous pouvons le remarquer, les scores de compréhensibilité sont globalement plus élevés dans la condition *a priori* que dans la condition *a posteriori*, avec, en particulier, moins d'énoncés ayant été associés avec la valeur la plus faible de l'échelle (1), et davantage d'énoncés ayant obtenu la valeur maximale (5). Afin de vérifier la significativité de cette différence de distribution, nous avons utilisé le test non-paramétrique de Wilcoxon. Les résultats ont mis au jour une différence significative entre les deux conditions, les scores de compréhensibilité relevés *a priori* étant plus élevés que ceux observés dans la condition *a posteriori* ($Z = -10,4$; $p < 0,001$, test unilatéral). Afin d'estimer la taille de l'effet, le r de Cohen a été calculé ; sa valeur est de 0,39, ce qui dénote une taille d'effet moyenne.

3.2 Accords inter-évaluateurs lors des évaluations *a priori* et *a posteriori*

Les accords inter-évaluateurs ont été calculés, pour chaque paire d'évaluateurs, sous la forme de coefficients de corrélation de Spearman. Sur l'ensemble des 40 paires d'évaluateurs, ces coefficients s'échelonnent de 0,16 à 0,98 dans la condition *a priori* (moyenne : 0,73) et de 0,25 à 1 dans la condition *a posteriori* (moyenne : 0,78). La Figure 3 présente les distributions des accords inter-évaluateurs dans ces deux conditions.

Nous pouvons observer que la valeur médiane des accords inter-évaluateurs est plus élevée dans la condition *a posteriori* (accord médian : 0,83) que dans la condition *a priori* (accord médian : 0,71).

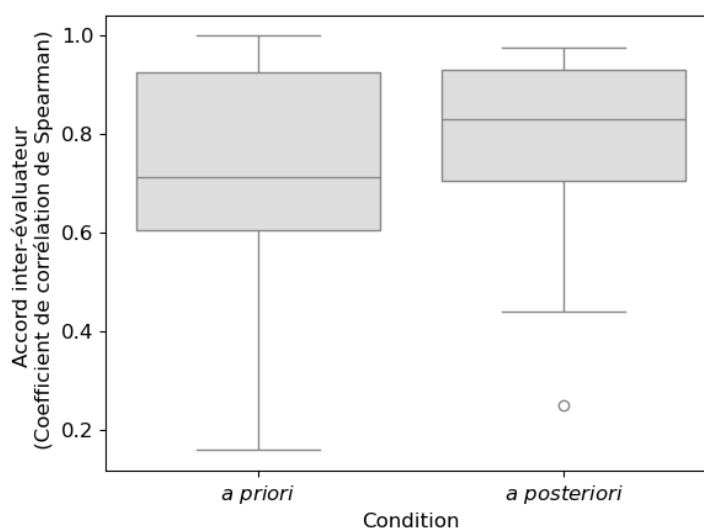


FIGURE 3 – Distribution des accords inter-évaluateurs pour les conditions *a priori* et *a posteriori*. Les barres d'erreur renvoient aux valeurs minimales et maximales, et les limites basses et hautes des boîtes aux 25^{ème} et 75^{ème} percentiles, respectivement. Les traits horizontaux à l'intérieur des boîtes représentent les valeurs médianes. Le cercle vide correspond à une donnée aberrante.

De même, l'étendue interquartile est plus faible dans la condition *a posteriori*, avec la moitié des valeurs comprises entre 0,70 et 0,93 (contre 0,6 et 0,92 pour la condition *a priori*). Afin de vérifier la significativité des différences d'accord observées entre les deux conditions, nous avons d'abord réalisé des tests de normalité de distribution à l'aide du test de Shapiro-Wilk. Les résultats ayant montré que les coefficients de corrélation ne sont pas normalement distribués ($p \leq 0,019$ dans les deux conditions), nous avons ensuite réalisé un test de Wilcoxon. Les résultats n'ont pas mis au jour de différence significative entre les accords inter-évaluateurs observés *a priori* et ceux relevés *a posteriori* ($Z = -1,3$; $p = 0,09$, test unilatéral).

4 Conclusion et discussion

Cette étude avait pour premier objectif de vérifier si l'absence de référence quant aux intentions de communication d'un locuteur pouvait constituer un biais lors de l'évaluation de la compréhensibilité de la parole d'énoncés correspondant à des traductions orales de phrases cibles. Plus précisément, nous supposons que, sans une telle référence, les évaluateurs auraient tendance à surestimer leur propre compréhension, dans la mesure où ils ne peuvent avoir conscience de divergences d'ordre sémantique entre le message perçu et le message que souhaitait transmettre le locuteur. Cette hypothèse a été confirmée par la mise au jour d'une différence significative entre les évaluations de compréhensibilité réalisées *a priori* et celles réalisées *a posteriori*, c'est-à-dire après la prise de connaissance d'une traduction de référence – les premières étant supérieures aux secondes. La taille (moyenne) de l'effet observé n'est pas négligeable, ce qui souligne d'autant plus l'enjeu que peut représenter un tel biais dans les études reposant sur des mesures de compréhensibilité obtenues par des méthodes d'évaluation *a priori*.

Il est néanmoins utile de rappeler que l'effet observé correspond à une tendance générale, observée à travers les différents énoncés produits par neuf apprenants germanophones, et dont la compréhensibilité a été évaluée par 80 auditeurs différents. Ceux-ci ont parfois également sous-évalué leur compréhensibilité lors de la condition *a priori*. Cette sous-évaluation représente toutefois une minorité de cas (18,8%) en comparaison des phénomènes de surévaluation (81,2%).

Notre seconde hypothèse était que les accords inter-évaluateurs seraient plus élevés dans la condition d'évaluation *a posteriori* que dans celle *a priori*, dans la mesure où disposer d'une référence commune devrait contribuer à l'harmonisation des évaluations. Malgré le calcul d'un nombre important d'accords (40), cette hypothèse n'a pas été confirmée par les statistiques inférentielles. Il est possible que le faible nombre de productions pour lesquels ces accords ont été calculés (9) n'ait pas permis d'obtenir des coefficients de corrélation suffisamment fiables pour mettre au jour une telle différence. Nous prévoyons, à terme, de mener une étude similaire incluant l'évaluation de 40 énoncés produits par 40 locuteurs japonophones de FLE (De Fino, 2024) afin de dépasser cette limite.

Remerciements

Ces travaux ont été financés par l'Agence Nationale de la Recherche dans le cadre du laboratoire commun ALAIA (ANR-18-LVC3-001), et par l'Association Nationale de la Recherche et de la Technologie *via* la thèse CIFRE de V. De Fino. Nous tenons à remercier le Pr. Gerndt pour nous avoir donné l'opportunité d'enregistrer des apprenants de FLE à l'université Ostfalia. Nous remercions également le Dr. Volmer et Mme Bizien pour l'apport de leur expertise lors de la création des énoncés à traduire, ainsi que les apprenants et évaluateurs qui ont participé à cette étude.

Références

- BERGERON A. & TROFIMOVICH P. (2017). Linguistic dimensions of accentedness and comprehensibility: Exploring task and listener effects in second language french. *Foreign Language Annals*, **50**(3), 547–566. DOI : [10.1111/flan.12285](https://doi.org/10.1111/flan.12285).
- CONSEIL DE L'EUROPE (2001). *Cadre Européen Commun de Référence pour les Langues : Apprendre, Enseigner, Évaluer*. Paris : Didier.
- CROWTHER D., TROFIMOVICH P., SAITO K. & ISAACS T. (2015). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, **49**(4), 814–837. DOI : [10.1002/tesq.203](https://doi.org/10.1002/tesq.203).
- CRUICKSHANKS K. J., WILEY T. L., TWEED T. S., KLEIN B. E., KLEIN R., MARES-PERLMAN J. A. & NONDAHL D. M. (1998). Prevalence of Hearing Loss in Older Adults in Beaver Dam, Wisconsin: The Epidemiology of Hearing Loss Study. *American Journal of Epidemiology*, **148**(9), 879–886. DOI : [10.1093/oxfordjournals.aje.a009713](https://doi.org/10.1093/oxfordjournals.aje.a009713).
- DE FINO V. (2024). *Caractérisation et mesure de la compréhension de la parole de locuteurs non natifs dans le cadre de l'apprentissage des langues*. Thèse de doctorat, Université Toulouse III Paul Sabatier.
- DERWING T. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, **59**(4), 547–567. DOI : [10.3138/cmlr.59.4.547](https://doi.org/10.3138/cmlr.59.4.547).
- DERWING T. M. & MUNRO M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, **39**(3), 379–397. DOI : [10.2307/3588486](https://doi.org/10.2307/3588486).
- DERWING T. M. & MUNRO M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, **42**(4), 476–490. DOI : [10.1017/S026144480800551X](https://doi.org/10.1017/S026144480800551X).
- FLEGE J. E. (1988). Factors affecting degree of perceived foreign accent in english sentences. *The Journal of the Acoustical Society of America*, **84**(1), 70–79. DOI : [10.1121/1.396876](https://doi.org/10.1121/1.396876).
- FONTAN L. (2012). *De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole pathologique en situation de communication*. Thèse de doctorat, Université de Toulouse. HAL : [tel-00797883](https://hal.archives-ouvertes.fr/tel-00797883).
- FONTAN L., GAILLARD P. & WOISARD V. (2013). Comprendre et agir : Les tests pragmatiques de compréhension de la parole et EloKanz. In R. SOCK, B. VAXELAIRE & C. FAUTH, Éd., *La voix et la parole perturbées*, p. 131–144. Mons: CIPA.
- HANSEN EDWARDS J. G., ZAMPINI M. L. & CUNNINGHAM C. (2018). The accentedness, comprehensibility, and intelligibility of Asian Englishes. *World Englishes*, **37**(4), 538–557. DOI : [10.1111/weng.12344](https://doi.org/10.1111/weng.12344).
- KANG O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, **38**(2), 301–315. DOI : [10.1016/j.system.2010.01.005](https://doi.org/10.1016/j.system.2010.01.005).
- KENNEDY S. & TROFIMOVICH P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, **64**(3), 459–489. DOI : [10.3138/cmlr.64.3.459](https://doi.org/10.3138/cmlr.64.3.459).
- LUDWIG A. & MORA J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation*, **3**(2), 167–198. DOI : [10.1075/jslp.3.2.01lud](https://doi.org/10.1075/jslp.3.2.01lud).
- MUNRO M. J. & DERWING T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, **44**(3), 316–327. DOI : [10.1017/S0261444811000103](https://doi.org/10.1017/S0261444811000103).

- NAGLE C. L. & HUENSCH A. (2020). Expanding the scope of L2 intelligibility research: Intelligibility, comprehensibility, and accentedness in L2 Spanish. *Journal of Second Language Pronunciation*, **6**(3), 329–351. DOI : [10.1075/bct.121.04nag](https://doi.org/10.1075/bct.121.04nag).
- SAITO K., MACMILLAN K., KACHLICKA M., KUNIHARA T. & MINEMATSU N. (2023). Automated assessment of second language comprehensibility: Review, training, validation, and generalization studies. *Studies in Second Language Acquisition*, **45**(1), 234–263. DOI : [10.1017/S0272263122000080](https://doi.org/10.1017/S0272263122000080).
- SAITO K., TROFIMOVICH P. & ISAACS T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, **38**(4), 439–462. DOI : [10.1093/applin/amv047](https://doi.org/10.1093/applin/amv047).
- SUZUKI S. & KORMOS J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, **42**(1), 143–167. DOI : [10.1017/S0272263119000421](https://doi.org/10.1017/S0272263119000421).
- SUZUKIDA Y. & SAITO K. (2021). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research*, **25**(3), 431–450. DOI : [10.1177/1362168819858246](https://doi.org/10.1177/1362168819858246).
- WOISARD V., ESPESSER R., GHIO A. & DUEZ D. (2013). De l'intelligibilité à la compréhensibilité de la parole, quelles mesures en pratique clinique ? *Revue de Laryngologie Otologie Rhinologie*, **134**(1), 27–33. HAL : [hal-01486715](https://hal.archives-ouvertes.fr/hal-01486715).