

Session 4

Atelier A.2.7

Nos items à l'accent québécois diffR-ils ? (7509)

*Dominique Casanova**, *Alhassane Aw**, *Mariam Bourras*** & *Marc Demeuse****

**Chambre de commerce et d'industrie de Paris Île-de-France, France*

***École nationale de statistique et d'analyse de l'information⁶⁸, France*

****Université de Mons, Belgique*

Mots-clés : évaluation en langue, accent, fonctionnement différentiel des items

Introduction

Dans un test de langue standardisé à vocation internationale, la question des accents utilisés dans les bandes son qui sont diffusées aux candidats est loin d'être anodine. Elle soulève des questions de désirabilité sociale, d'équité et de standardisation. L'introduction ou non de messages oraux comportant des accents particuliers dépend bien entendu de l'objectif du test, de son usage par les prescripteurs et du ou des publics ciblés. Les choses se compliquent lorsqu'un même test est utilisé pour des enjeux différents et/ou par des prescripteurs de pays différents ayant une langue en partage. Des tensions apparaissent alors, compte-tenu de la multiplicité des publics et des usages du test.

Cette question se pose notamment pour le Test d'évaluation de français - TEF (Noël-Jothy & Sampsonis, 2006 ; 72-74, Pons & Karcher, 2006). Si le TEF a été conçu à l'origine pour répondre à une demande et aux besoins d'écoles de commerce françaises qui accueillaient de nombreux étudiants étrangers, il est aujourd'hui fortement utilisé dans le cadre de démarches d'immigration au Canada ou au Québec.

Jusqu'à récemment, le TEF comportait exclusivement des bandes son à l'accent réputé « neutre », comme celui souvent proposé dans les méthodes d'enseignement/apprentissage éditées à France à l'attention d'un public allophone. Cet accent « standard » du TEF, qui s'inspire du « français des médias » pour les monologues ou interviews, est cependant souvent perçu comme un accent « français » par la société civile au Canada et au Québec et par les candidats qui passent le test sur place. Ce sentiment était renforcé par le manque de référents culturels nord-américains (références monétaires, géographiques, sportives...) dans les textes composant le test.

La Chambre de commerce et d'industrie de région Paris Ile-de-France a aisément pu diversifier les référents culturels francophones, les questions du test ne portant pas sur ces spécificités régionales qui « habillent » le texte. Elle a en revanche souhaité limiter dans un premier temps le recours à des accents variés, pour mesurer l'effet qu'ils pouvaient avoir sur la compréhension par des candidats issus de zones géographiques différentes. L'introduction d'un accent même modéré est en effet susceptible de modifier la difficulté de la compréhension d'un message (Ockey & French, 2014), notamment pour les candidats ayant un niveau de compétence peu élevé.

La question de l'impact de l'accent sur la compréhension de messages en langue française par des allophones a jusqu'à présent été peu investiguée. Les cadres auxquels le TEF fait référence, que ce soit le Cadre européen commun de référence pour les langues (Conseil de l'Europe, 2000) ou les Niveaux

⁶⁸ Cette communication fait suite à un stage de deuxième année réalisé par Mariam Bourras au sein de la Chambre de commerce et d'industrie de région Paris Ile-de-France.

de compétences linguistiques Canadiens (Centre des niveaux de compétence linguistique canadiens, 2012), sont eux-mêmes assez peu diserts sur la question des accents. On trouve davantage de références dans la littérature anglophone où la question s'étend à l'usage d'accents de locuteur non-natifs anglophones dans des tests de compréhension en langue anglaise (Abeywickrama, P., 2013 ; Harding, L. 2011).

Dans le cadre de cette étude, nous avons mis à profit la présence de quelques messages à l'accent québécois dans les derniers questionnaires de l'épreuve de compréhension orale du TEF. Les données des épreuves des dernières années ont donc été analysées pour mesurer l'impact éventuel de la différence d'accent (québécois versus « standard ») sur les performances des candidats provenant des différentes régions du monde et passant le test dans l'objectif d'étudier en France ou d'immigrer ou de rester durablement au Canada ou au Québec. L'hypothèse émise par les concepteurs du test était que l'introduction d'un accent québécois modéré dans les messages oraux de certains items ne favorise pas la réussite à ces items des candidats passant le test sur le territoire canadien (majoritairement à Montréal).

L'objet de l'étude est d'analyser les réponses aux items de deux catégories principales de population (ceux qui ont passé le test au Canada et ceux qui l'ont passé en France), pour voir si les items à l'accent québécois ont tendance à favoriser certains groupes de population par rapport à d'autres, étant donné le résultat global à l'épreuve de ces candidats. Il s'agit donc de détecter un éventuel fonctionnement différentiel (FDI) de ces items (Camilli et Shepard, 1994), qui pourrait apparaître en faveur des candidats passant le TEF au Canada. Nous avons pour cela mobilisé différentes méthodes d'analyse du fonctionnement différentiel des items telles qu'implémentées dans la librairie difR (Magis et al., 2010).

La question de l'impact de l'accent sur la compréhension en langue étrangère dans la littérature scientifique

La question même de ce que constitue un accent est loin d'être triviale. Selon Raymond Renard (1979), « acoustiquement, l'accent est lié essentiellement à la variation d'intensité de la voix, bien que la durée, la hauteur et/ou le timbre puissent également jouer un rôle compensatoire ». D'autres chercheurs insistent sur sa dimension sociolinguistique (Fries & Deprez, 2003) et le fait que l'accent puisse être défini comme la façon dont le langage parlé d'un locuteur diffère de la variété locale du groupe d'individus qui l'écoute et de l'impact de cette différence sur les différents interlocuteurs (Derwing & Munro, 2009).

Le Cadre européen commun de référence pour les langues ne propose pas de définition de ce qui constitue l'accent. Il y fait cependant référence dans certains descripteurs de certaines échelles. En compréhension générale de l'oral, il introduit la notion d'accent au niveau B1+ (sur une échelle allant de A1 – utilisateur élémentaire, à C2 – utilisateur expérimenté), en indiquant qu'à ce niveau l'apprenant « Peut comprendre une information factuelle directe sur des sujets de la vie quotidienne ou relatifs au travail en reconnaissant les messages généraux et les points de détail, **à condition que** l'articulation soit claire **et l'accent courant** » et au niveau C1 qu'il « Peut suivre une intervention d'une certaine longueur sur des sujets abstraits ou complexes même hors de son domaine **mais peut avoir besoin de faire confirmer quelques détails, notamment si l'accent n'est pas familier** ».

De même la question de l'accent, dans les « profils de compétence pour la compréhension de l'oral » des Niveaux de compétences linguistiques Canadiens (Immigration, Réfugiés et Citoyenneté Canada, 2012), n'est évoquée qu'à partir du niveau 8 (soit à la fin du stade II – Intermédiaire, de l'échelle à 12 niveaux des NCLC). On peut y lire « Suit des conversations rapides entre locuteurs natifs **qui n'ont pas un accent prononcé** et n'emploient pas de régionalismes », et ce n'est qu'à partir du niveau 9 (soit au

début du stade III – Avancé) que la performance globale mentionne « **Peut comprendre certains accents régionaux** et des régionalismes ».

Selon ces référentiels, la présence d'accents est donc susceptible d'être un frein à la compréhension de textes oraux pour les locuteurs de niveau élémentaire ou intermédiaire d'une langue, ce qui peut encourager les concepteurs de tests à réserver l'utilisation d'accents variés aux questions de niveau le plus élevé. Mais ce qui constitue un accent familier pour un candidat diffère selon qu'il réside dans un pays francophone d'Afrique, d'Asie, d'Europe ou d'Amérique du Nord, ou dans un pays non francophone. Réserver l'utilisation d'accents variés aux questions de niveau le plus élevé ne risque-t-il au contraire pas d'être inéquitable envers les candidats moins familiers de l'accent « standard » du test ? Dans un monde globalisé, la capacité à comprendre des locuteurs à l'accent varié ne doit-elle pas faire partie du construit d'un test de compréhension orale ?

La question fait débat dans le monde anglophone et s'étant même au-delà, certains chercheurs s'intéressant à l'usage d'accents non-anglophones dans des tests de compréhension en langue anglaise (Harding, L. 2011). Plusieurs études ont été menées en ce sens et conduisent à des résultats qui peuvent paraître contradictoires mais sont peut-être dus aux contextes expérimentaux. Ainsi, alors qu'Abeywickrama (2013) rapporte l'absence d'effet de l'accent utilisé (américain, chinois, coréen ou sri-lankais) sur les résultats à un questionnaire de compréhension délivré à des étudiants brésiliens, coréens et sri-lankais, Major et al. (2002) montrent, dans leur étude, que les résultats de candidats natifs et non-natifs se sont avérés plus faibles lorsque ces candidats ont été exposés à des locuteurs non-natifs de langue anglaise.

Gary Ockey et Robert French (2014) ont proposé une méthode pour établir une relation éventuelle entre l'intensité d'un accent et la compréhension du message, d'une part, et entre la familiarité avec un « type » d'accent et la compréhension du message, d'autre part. Ils ont construit une échelle d'intensité de l'accent (« Strength of Accent Scale ») qu'ils ont utilisée auprès d'une centaine d'étudiants et d'enseignants d'établissements américains pour classer 20 locuteurs anglophones selon l'intensité de leur accent. Ils ont ensuite retenu 9 locuteurs (un américain, quatre anglais et quatre australiens) dont l'accent était d'intensité variable pour proposer 9 versions parallèles d'un même texte oral représentatif d'une section du TOEFL iBT (monologue de 686 mots portant sur un thème de sciences naturelles), associées aux 6 mêmes questions de compréhension. Les candidats d'une cohorte entière du TOEFL iBT (21 726 candidats provenant de 148 pays différents) se sont vu attribuer aléatoirement une de ces versions et ont répondu (pour un quart d'entre eux) à un questionnaire portant sur leur familiarité avec l'accent utilisé. Cette étude a permis de montrer que les résultats étaient, à l'exception d'un locuteur, d'autant plus faibles que l'intensité de l'accent du locuteur (qualifié par des personnes résidant aux États-Unis) était élevée. Elle a aussi mis en évidence un lien entre la familiarité avec l'accent du locuteur et la performance aux questions de compréhension portant sur ce texte. Quoique limités, ces effets sont apparus alors que seuls des accents modérés ont été pris en considération dans l'étude.

L'analyse du fonctionnement différentiel des items à l'accent québécois du TEF

Si la présence d'un accent même modéré est susceptible de modifier la compréhension d'un texte oral par les candidats selon leur familiarité avec le type d'accent en question, alors les effets de la présence de quelques items à l'accent québécois devraient être perceptibles en comparant les réponses aux items des candidats passant ce test au Canada à ceux des candidats passant ce test en France, dans un autre pays francophone ou dans un pays non francophone. A performance égale au test, les candidats familiers avec l'accent québécois (approximés dans cette étude en considérant les candidats ayant

passé le test sur le territoire canadien) devraient mieux réussir les items à l'accent québécois que les autres.

Ainsi, en analysant la présence d'un éventuel fonctionnement différentiel des items selon la variable « groupe d'appartenance », en prenant comme groupe focal les candidats ayant passé le test au Canada et comme groupe de référence les candidats ayant passé le test en France, il devrait être possible de mettre en évidence l'effet de la familiarité avec l'accent québécois sur la réussite aux items de compréhension de textes énoncés avec cet accent. Nous définissons ici avec Bertrand et Blais (2004) le fonctionnement différentiel d'un item le cas où « deux sujets d'habileté égale mais appartenant à des groupes distincts ont une probabilité différente de réussir l'item ».

Il existe cependant un nombre important de méthodes différentes de détection du fonctionnement différentiel des items, dont un bon nombre sont décrites dans Bertrand et Blais (2004). 9 de ces méthodes sont implémentées dans la librairie difR (Magis et al., 2010). Ces différentes méthodes ne donnent cependant pas toujours les mêmes résultats et il est d'usage d'en convoquer plusieurs pour parvenir à une meilleure interprétation des données.

Pour notre étude, nous avons retenu les méthodes suivantes :

- La méthode Mantel-Haenszel (Holland et Thayer, 1986), méthode non paramétrique parmi les plus répandues pour la détection de FDI uniformes (cas où une différence probabilité de réussite dans la même direction peut être constatée indépendamment du score de référence des individus) ;
- Une méthode basée sur la régression logistique (Clauser et Mazor, 1998), qui permet de détecter des items présentant un DFI uniforme ou non uniforme (cas où le signe de la différence des probabilités de réussite peut varier selon le score de référence des individus) ;
- Une méthode basée sur la théorie de réponse aux items. Parmi les méthodes paramétriques proposées dans la librairie difR, nous avons choisi la méthode de Lord (Raju et al., 1995), en l'appliquant à un modèle à 1 paramètre, les méthodes alternatives conduisant à des erreurs ou à l'absence de convergence des calculs.

La librairie difR propose par ailleurs des règles d'interprétation comparables pour l'interprétation des résultats. Elle fournit une information sur la significativité des statistiques, puis classe l'effet selon sa taille en référence aux catégories A, B et C définies initialement par ETS pour la méthode de Mantel-Haenszel (Zwick and Ercikan, 1989). Lorsque la statistique est significative, la catégorie A correspond à un FDI négligeable, la catégorie B à un FDI modéré et la catégorie C à un FDI sévère (Bertrand et Blais, 2004).

Nous avons considéré les résultats à 4 versions récentes du Test d'évaluation de français (intitulées ci-après version 1 à 4) comportant chacune 3 ou 4 questions associées à des messages à l'accent québécois (sur 50 questions se rapportant à des messages allant au-delà d'une phrase courte). Le tableau 1 présente la répartition géographique des effectifs des candidats à ces tests. Si pour la version 1 et 2 les effectifs entre les groupes Canada et France sont comparables, on constate un déséquilibre dans le cas des versions 3 et 4.

Tableau 1 : répartition géographique des candidats selon le lieu de passation de l'épreuve, en nombre d'individus

	Canada	France	Autre pays francophone	Reste du monde
Version 1	720	715	1703	1058
Version 2	589	520	1162	635
Version 3	1032	499	1373	351
Version 4	1640	778	2203	891

Le TEF ayant plusieurs usages, nous n'avons considéré dans cette étude que les candidats inscrits dans un processus de mobilité (pour des études en France, pour l'accès au territoire ou la résidence au Canada ou au Québec). Ces différents publics sont relativement comparables tant d'un point de vue sociologique qu'en termes de niveau de langue (un niveau au moins égal à B2 étant en général requis pour leur mobilité). La figure 1 représente la densité du score de compréhension orale des candidats au TEF de l'année 2017 selon le lieu de passation du test.

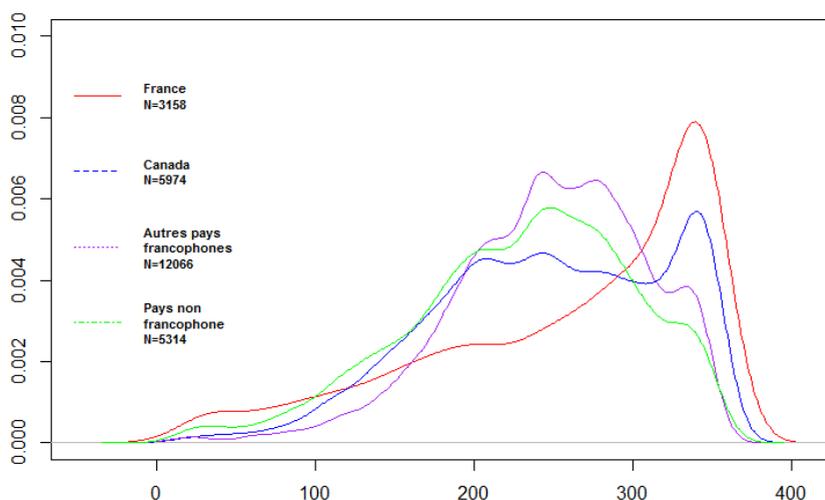


Figure 1 : répartition par score des candidats au TEF ayant un projet de mobilité selon le pays de passation des épreuves.

Résultats

Une analyse du fonctionnement différentiel des items a été menée pour chacune des quatre versions de l'épreuve de compréhension orale, en considérant comme groupe de référence les candidats ayant passé le test en France et comme groupe focal ceux qui l'ont passé au Canada. Le tableau 2 présente, pour chacune des versions de l'épreuve et pour chacune des méthodes utilisées, le nombre d'items pour lesquels le test est significatif, indépendamment de la taille de l'effet.

Tableau 2 : nombre de cas significatifs par version, pour chaque méthode

	Mantel-Haenszel	Régression logistique	Lord (1PL)
Version 1	10	12	8
Version 2	12	11	14
Version 3	11	25	18
Version 4	20	37	22

Les résultats montrent, pour la plupart des méthodes, que le nombre d'items détectés comme présentant un FDI va croissant avec le déséquilibre entre les effectifs des deux groupes. Cela est particulièrement visible pour la méthode de régression logistique, qui est réputée sensible au nombre de sujets dans chacun des groupes (Bertrand et Blais, 2004). Le tableau 3 présente le nombre de ces items qui appartiennent à la catégorie B (taille d'effet modérée) ou C (taille d'effet sévère) et sont donc susceptibles d'être problématiques.

Tableau 3 : nombre d'items présentant un DIF modéré ou sévère selon chacune des méthodes utilisées

	Mantel-Haenszel	Régression logistique	Lord (1PL)
Version 1	6	0	8
Version 2	8	1	14
Version 3	8	0	13
Version 4	11	1	12

Les résultats montrent cette fois-ci, dans le cas de la méthode de régression logistique, que la quasi-totalité des items dont la statistique est significative sont associés à une taille de catégorie A (négligeable), même lorsqu'une approche libérale est retenue pour la définition des seuils⁶⁹. Cette méthode étant sensible au nombre de sujets dans chacun des groupes, Bertrand et Blais (2004) proposent de procéder au déploiement du diagramme en boîte et moustaches des valeurs du khi-carré, puis de comparer la valeur du khi-carré des items dont la valeur statistique est significative à l'écart interquartile, pour les répartir dans les catégories A, B et C. Cela ne conduit pas à une augmentation notable du nombre d'items des catégories B et C. Il semble donc que la méthode de régression logistique, du moins tel qu'implémentée dans la librairie difR, ne soit pas adaptée à notre étude.

Les méthodes de Mantel-Haenszel et de Lord donnent des résultats plutôt concordants, mais force est de constater que très peu des items ainsi étiquetés sont des items à l'accent québécois. En effet, sur les 10 items à l'accent québécois répartis dans les différentes versions (3 de ces items étaient présents dans les versions 1 et 2), seuls 3 ont été détectés comme présentant un fonctionnement différentiel modéré ou sévère par l'une ou les deux méthodes. Le tableau 4 rapporte les résultats des traitements pour les items à l'accent québécois pour chacune de ces deux méthodes.

⁶⁹ Il existe en effet deux propositions très différentes pour délimiter les catégories A, B et C pour la méthode de régression logistique (celle formulée par Jodoin et Gierl, 2001 et celle proposée par Zumbo et Thomas, 1999).

Tableau 4 : nombre d'items présentant un DIF modéré ou sévère selon chacune des méthodes utilisées

		Mantel-Haenszel	Lord (1PL)
U2347-1	Version 1	n.s.	n.s.
	Version 2	n.s.	n.s.
U2347-2	Version 1	n.s.	n.s.
	Version 2	n.s.	n.s.
U2754	Version 1	B***	B***
	Version 2	B**	n.s.
U729-1	Version 3	B**	C***
U729-2	Version 3	n.s.	n.s.
U4153.4	Version 3	B*	B**
U4046-1	Version 4	n.s.	n.s.
U4046-2	Version 4	n.s.	n.s.
U1846	Version 4	n.s.	n.s.

n.s. Valeur non significative

* Valeur significative à $p < .05$

** Valeur significative à $p < .01$

*** Valeur significative à $p < .001$

L'item U2754, utilisé dans les versions 1 et 2, présente un fonctionnement différentiel modéré très significatif dans la version 1, mais seule la méthode de Mantel Haenszel détecte un FDI significatif pour la version 2. L'item 4153 présente également un fonctionnement différentiel selon les deux méthodes. Le cas des items U729-1 et U729-2 est singulier. Ces deux items se rapportent en effet à une même bande son, le premier item ayant pour objectif de déterminer la nature du message (amicale, familiale, professionnelle ou publicitaire) et le second portant sur l'information transmise. Il semblerait que la présence d'un accent québécois n'ait pas favorisé la compréhension du message par les candidats d'un des deux groupes mais davantage la compréhension de la relation sociale entre les interlocuteurs.

Pour faciliter l'interprétation du fonctionnement différentiel, on peut tracer, sur un même graphe, le pourcentage de réussite des candidats de chaque groupe selon le score total. La figure 2 présente cette information pour l'item U729-1. Compte-tenu des effectifs somme-toute limités, le pourcentage de réussite a été déterminé non pas pour chacun des scores possibles, mais par intervalle de 5 points (et pour les intervalles regroupant au moins 10 individus). On constate que, en dépit de l'accent québécois, ce sont les candidats ayant passé le test en France qui, à score total équivalent, ont mieux réussi cet item.

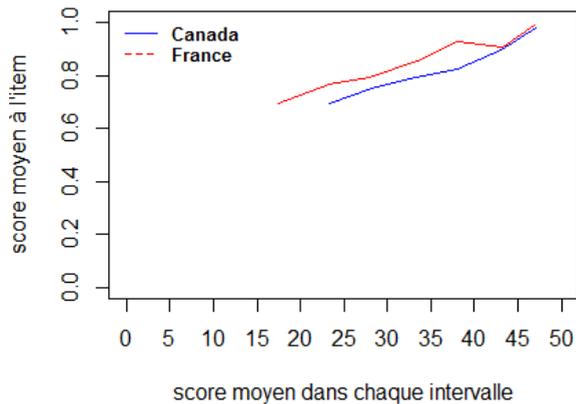


Figure 2 : pourcentage de réussite de l'item U729-1 par les candidats du Canada et de France selon leur score total au test

On peut s'interroger sur la présence éventuelle d'un trait culturel dans le ton employé ou la formule d'adresse, qui aurait tendance à favoriser les candidats résidant en France, mais on s'attendrait alors à retrouver ce résultat en comparant les résultats obtenus par les autres groupes de population. La situation n'est cependant pas si nette au regard de la figure 3. Si les candidats ayant passé le test dans d'autres pays francophones ont clairement sous-performé à cet item, ce n'est pas le cas des candidats ayant passé le test dans un pays non francophone. Faut-il en déduire une plus proximité culturelle entre la France et ces pays non francophones très divers qu'avec le Canada ?

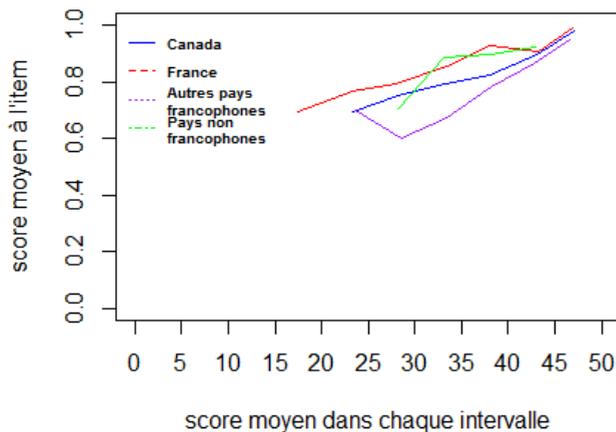


Figure 3 : pourcentage de réussite de l'item U729-1 pour chacune des catégories de population

Le cas le plus net concerne l'item U2754 (Cf. figure 4), qui semble avantager les candidats résidant au Canada, tant pour la version 1 que pour la version 2. Pourtant, comme le montre la figure 5, les probabilités de réussite des autres catégories de population sont très proches de celles qu'on constate pour les candidats résidant au Canada (Cf. figure 4). Ce seraient donc surtout les candidats passant le test en France qui seraient désavantagés par ce message à l'accent québécois (mais pas par les précédents), contrairement à l'ensemble des candidats.

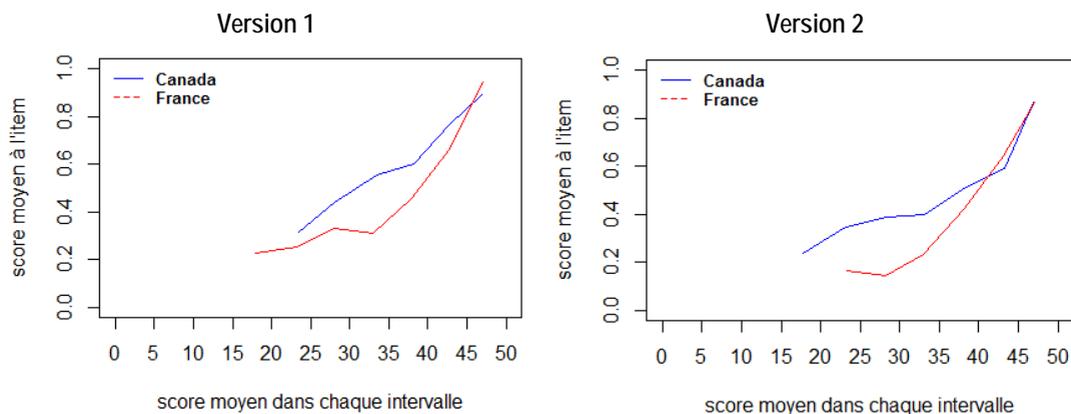


Figure 4 : réussite comparée de l'item U2754 par les candidats ayant passé le test au Canada ou en France, pour les versions 1 et 2 de l'épreuve de compréhension orale

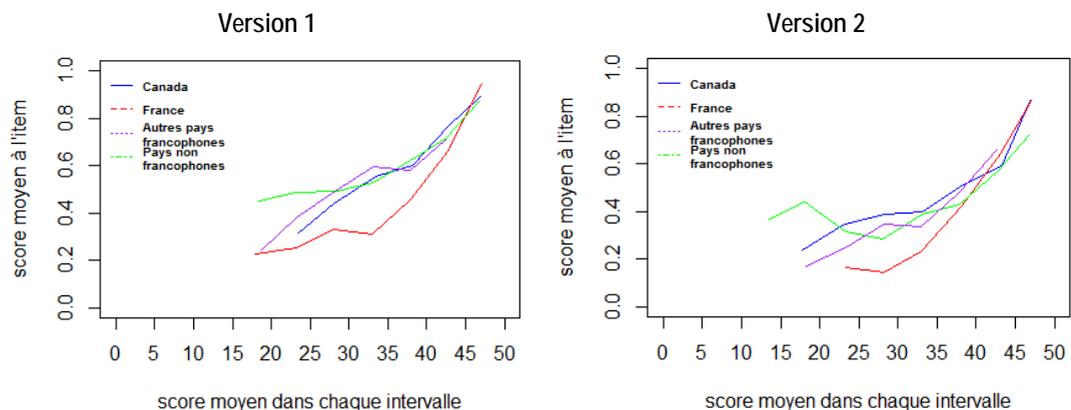


Figure 5 : réussite comparée de l'item U2754 par les différentes catégories de candidats

Enfin, l'item U4153.4 présentait un FDI modéré selon les deux méthodes. La figure 6 confirme un léger avantage pour les candidats résidant au Canada par rapport aux candidats résidant en France, mais pas d'avantage net par rapport aux candidats résidant dans un autre pays francophone. Pour les candidats résidant dans un pays non francophone, ce sont les plus faibles qui semblent être comparativement désavantagés par cet item.

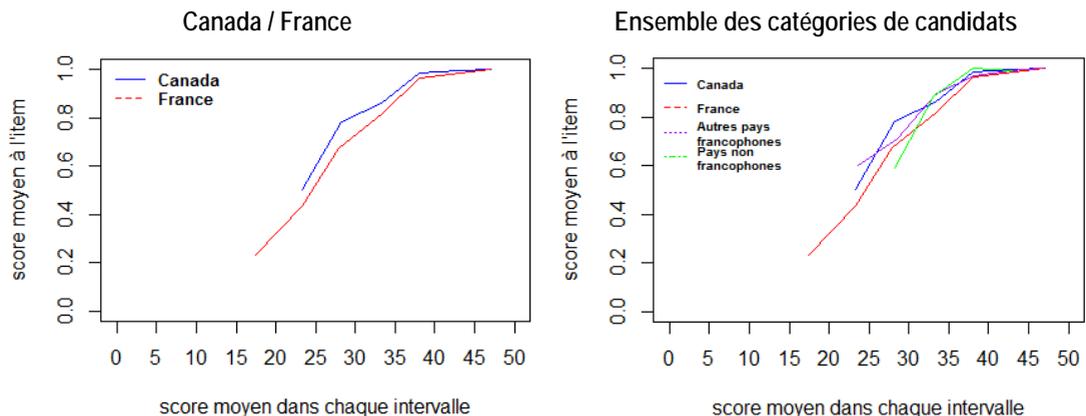


Figure 6 : réussite comparée de l'item U4153.4 par les différentes catégories de candidats

Quand on considère les autres items pour lesquels les méthodes détectent un fonctionnement différentiel, cela rajoute à la confusion, puisque certains de ces items semblent favoriser les candidats résidant au Canada alors que c'est l'accent « standard » du test qui est utilisé. D'autres items à l'accent standard sont réussis de manière similaire pour des candidats de même niveau général résidant au Canada, en France ou dans un pays non francophone mais clairement mieux réussis ou davantage échoués par les candidats résidant dans un autre pays francophone.

Discussion

Au vu de ces résultats contrastés, il semble que l'accent ne joue qu'un rôle limité dans les différences de compréhension d'un message du TEF selon que les candidats résident au Canada ou en France. Sur les dix items portant sur un message à l'accent québécois, seuls 3 présentent statistiquement un fonctionnement différentiel, dont un semblant favoriser les candidats passant le test en France. Par ailleurs, certains items à l'accent « standard » du TEF semblent favoriser les candidats passant le test au Canada. Il y a donc probablement d'autres facteurs intervenant dans la compréhension orale d'un message qui contribuent au fonctionnement différentiel des items et la prudence est de mise dans l'interprétation des résultats d'une analyse du fonctionnement différentiel des items.

Plusieurs facteurs peuvent expliquer ces résultats, qui s'accordent peu avec la littérature de référence. D'une part l'accent utilisé était modéré et restait dans l'esprit de l'accent « standard » du TEF, prenant pour référence l'accent des médias internationaux. D'autre part, ce sont principalement des monologues courts qui ont été utilisés dans ces versions du test. Des résultats différents pourraient apparaître sur de longues interviews. Enfin, la population qui passe le test a en général un niveau B1+ ou supérieur : l'effet potentiel de l'accent sur des candidats de niveau élémentaire ne peut donc pas être analysé.

Cela est rassurant au sens que le standard utilisé dans le TEF en termes d'accent n'a apparemment pas de conséquences fâcheuses sur la compréhension des messages oraux par les candidats qui seraient liées au pays de résidence. Peut-être est-ce simplement dû au fait que, compte-tenu des enjeux pour les candidats, ces derniers se familiarisent avec l'accent utilisé lors de leur préparation au test. Cependant ce standard gagnerait sans doute à évoluer pour être plus représentatif de la diversité francophone, tout en conservant sa référence aux médias internationaux par souci de compréhensibilité. Compte-tenu du fait que la grande majorité des candidats ont pour objectif de s'installer durablement au Canada, au Québec ou en France, il ne semble pas absurde d'intégrer une part significative d'items à

l'accent québécois dans le test. Cela permettrait d'accroître la validité d'usage des résultats du test dans les procédures d'immigration canadiennes et québécoises par une prise en considération de la capacité à comprendre des messages oraux énoncés avec un accent canadien modéré, susceptible de favoriser l'insertion sociale. Au-delà de ces deux accents, le test gagnerait à représenter plus largement la francophonie en introduisant une plus grande variété de référents culturels, tant à l'écrit qu'à l'oral. Les candidats au TEF sont répartis dans plus d'une centaine de pays et peuvent être en contact avec différentes variétés locales de la langue française, qui en font sa richesse. Encore faut-il être en mesure de proposer un échantillonnage satisfaisant de cette diversité dans le test.

Conclusion

Dans la présente étude, nous avons cherché à mesurer l'impact éventuel de la présence de messages audio présentant un accent québécois modéré sur la compréhension orale de candidats selon leur lieu de résidence (Canada ou France). Nous avons pour cela mobilisé différentes méthodes de détection du fonctionnement différentiel des items.

Les résultats contrastés semblent montrer que, dans les conditions d'enregistrement du Test d'évaluation de français (TEF) et pour le public concerné, la familiarité avec l'accent ne constitue pas un déterminant majeur de la compréhension orale des textes. Des études s'appuyant sur des principes différents mériteraient d'être entreprises pour confirmer cette impression.

Une première approche serait de présumer l'existence d'une deuxième dimension dans les données liée à la présence des items à l'accent québécois, puis de confirmer ou d'infirmer la présence de cette dimension en appliquant un modèle de réponses aux items à deux dimensions (où seuls les items à l'accent québécois seraient concernés par cette deuxième dimension). Une autre façon d'analyser la dimensionnalité serait d'appliquer la méthode non paramétrique de Stout (1987, 1990) au moyen du logiciel DIMTEST.

Une seconde approche serait de proposer un devis expérimental où des messages seraient enregistrés avec des accents différents et où ces différentes versions seraient proposées de façon aléatoire aux candidats. La possibilité de mise en œuvre d'une telle expérimentation reste à confirmer.

Quoiqu'il en soit, tout être humain a sa façon propre de prononcer, d'articuler et de marquer phonétiquement ses propos. La capacité à traiter un message oral pour le comprendre implique de s'adapter aux particularités phonétiques de l'interlocuteur. Il semble donc pertinent d'inclure une variété d'accents dans un test de langue, plutôt que de promouvoir un accent particulier au motif qu'il constituerait le « standard » du test. C'est probablement davantage les choix retenus pour l'échantillonnage de ces accents qui doivent être questionnés, au regard de la vocation du test et des usages qui sont faits des résultats.

Bibliographie

- Abeywickrama, P. (2013). Why Not Non-native Varieties of English as Listening Comprehension Test Input ? *RELC Journal* 44(1), p. 59-74.
- Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology : Human Perception and Performance*, 35(2), p. 520-529.
- Bertrand, B. et Blais, J. G. (2004). *Modèles de mesure. L'apport de la théorie des réponses aux items*. Sainte Foy (Québec) : Presses de l'Université du Québec.
- Camilli, G. et Shepard, L.A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA : Sage.

- Clauser, B.E. et Mazor, K.M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement : Issues and Practices, printemps*, p. 31-44.
- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues*. Paris : Didier.
- Derwing, T.M. et Munro, M.J. (2009). Putting accent in its place : Rethinking obstacles to communication. *Language Teaching* 42(4), p. 476-490.
- Fries, S. et Deprez, C. (2003). L'accent étranger : identification et traitement social en France et aux États-Unis. *Français : variations, représentations, pratiques* (éds. Jacqueline Billiez & Didier de Robillard), *Cahiers du français contemporain*, n°8, Lyon : ENS Éditions.
- Harding, L. (2011). *Accent and Listening Assessment : A Validation of the Use of Speakers with L2 Accents on an Academic English Listening Test*. Franckfurt am Main : Peter Lang.
- Holland, P.W. et Thayer, D.T. (1986). Conditional association and unidimensionality assumption in monotone latent variable models. *The Annals of Statistics*, 14, p. 1523-1543.
- Immigration, Réfugiés et Citoyenneté Canada (2012). Niveaux de compétence linguistiques canadiens : français langue seconde pour adultes. Ottawa : Centre des niveaux de compétence linguistique canadiens.
- Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42, p.847-862.
- Major, R. C., Fitzmaurice, S. F., Bunta, F., Balasubramanian, C. (2002). The Effects of Nonnative Accents on Listening Comprehension : Implications for ESL Assessment. *TESOL Quarterly Vol. 36*(2), p. 173-190.
- Noël-Jothy & Sampsonis (2006). *Certifications et outils d'évaluation en FLE*. Paris : Hachette.
- Ockey & French (2014). From One to Multiple Accents on a Test of L2 Listening Comprehension. *English Publications 83*. Oxford University Press.
- Pons, S. et Karcher, G. (2006). *TEF 250 activités*. Paris : Clé International.
- Raju, N.S., van der Linden, W.J. et Fleer, P.F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19(4), p. 353-368.
- Renard, R. (1979). La méthode verbo-tonale de correction phonétique. Centre International de Phonétique Appliquée – Mons. Paris : Didier Érudition.
- Zwick, R. et Ercikan, K. (1989). Analysis of Differential Item Functioning in the NAEP History Assessment. *Journal of Educational Measurement, Vol. 26* (1), p. 55-66.

Faut-il lire, et que faut-il lire, pour mieux savoir lire ? (7617)

Monique Reichert, Charlotte Kraemer, Salvador Rivas, Rachel Wollschlaeger & Sonja Ugen
 Université du Luxembourg, Luxembourg

Mots-clés : compétences en lecture, habitudes de lecture, arrière-fond socio-culturel

En matière d'éducation en général, et plus précisément dans le contexte de l'acquisition de compétences et de connaissances scolaires, on ne cesse de souligner que dans beaucoup de pays il existe des écarts importants entre des élèves en fonction de leur arrière-fond socio-économique (SES), de leur langue maternelle, de leur statut de migration, ou du genre. Sachant que la formation scolaire peut influencer de manière cruciale sur le parcours de vie ultérieur, il importe de clarifier quelles pourraient être d'autres caractéristiques qui interagiraient avec les variables pré-mentionnées, qui aideraient, premièrement, à mieux saisir le pourquoi de ces différences, et, deuxièmement, à élucider quelles mesures de soutien pourraient être adoptées. Si l'on focalise sur les compétences en lecture, les chercheurs recourent parfois aux habitudes de lecture des élèves en supposant que la fréquence de lecture puisse avoir un