Snap judgement : l'influence de l'origine ethnique, réelle ou imaginée, sur les évaluations des compétences en langue étrangère

Claire Wells

Université Stendhal - Grenoble III
UFR de Sciences du langage

Mémoire de recherche

préparé en vue de l'obtention du

Master en Sciences du Langage

Présenté par :

Claire WELLS

Sous la direction de Jean-Pierre CHEVROT
Date de soutenance : 9 septembre 2013

SNAP JUDGEMENT :

L'INFLUENCE DE L'ORIGINE ETHNIQUE, RÉELE OU IMAGINÉE, SUR LES ÉVALUATIONS DES COMPÉTENCES EN LANGUE ÉTRANGÈRE

Université Stendhal -Grenoble 3
1180, Avenue centrale, BP 25, 38040 Grenoble cedex 9

UNIVERSITE Stendhal GRENOBLE 3

Déclaration anti-plagiat

Document à scanner après signature
et à joindre au mémoire électronique

DECLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : Wells PRENOM : Claire

DATE : Sept 5/13 SIGNATURE : [redacted]

Remerciements

Je tiens tout d'abord à remercier Jean-Pierre Chevrot qui a accepté de diriger mon mémoire sans m'avoir rencontrée en personne. Ses conseils, son encouragement et son investissement constant, m'ont beaucoup apporté et m'ont permis de réaliser ce travail.

Je souhaite remercier tous les étudiants qui m'ont aidée avec ce travail. Les trois "étudiantes étrangères", Ruba, Yuhui et Karen, ainsi que tous les étudiants francophones qui ont fait les questionnaires. Je remercie également Mme Ounoughi, Mme Cannard et M. Chauvin, de m'avoir laissé passer le protocole dans leurs cours.

Je souhaite remercier en dernier lieu mon entourage. Un grand merci à Christine, ma relectrice ainsi qu'à Wendy et Monique pour leurs conseils et leurs encouragements. Finally, I would like to thank my husband, Seth, for his constant support and his wisdom regarding statistics and the writing process.

Table des matières

Table des tableaux

Table des figures

INTRODUCTION

La rédactrice de ce mémoire, anglophone professeure de français, réfléchit souvent à l'évaluation des compétences en langue étrangère. Ayant passé le DELF B2 (diplôme de la langue française) en tant que candidat puis désormais examinatrice, nous sommes pleinement consciente de la complexité de l'évaluation des compétences en langue orale à ce niveau de DELF. Les critères sur lesquels les décisions sont fondées ne peuvent pas être trop précis et sont donc ouvert à des interprétations différentes. Qu'est-ce qui fait la différence, par exemple, entre un score de 2/3 et de 3/3 pour le critère de : "possède, malgré son accent, une prononciation intelligible" (Council of Europe, 2001)?

Historiquement, ces évaluations ont été vues comme le résultat de considérations conscientes et bien réfléchies (Bargh & Chartrand, 1999). Ainsi, les juges de ces examens aux conséquences importantes suivent une formation rigoureuse pour assurer un certain degré de *fiabilité inter-juge*[1], terme qui désigne les différences de jugements entre plusieurs juges d'un seul postulant (McNamara, 1996). Pourtant, de nouvelles théories sur les processus cognitifs, provenant des sciences cognitives et de la psychologie, remettent en question l'influence du conscient sur les jugements (entre autres Ambady & Rosenthal, 1992; Bargh & Chartrand, 1999; Campbell-Kibler, 2012; Niedzielski, 1999). En effet, il semble que la perception de la parole, et par conséquent les conclusions tirées des perceptions, peuvent être influencées de façon importante par plusieurs facteurs : ce que le juge voit (McGurk & Macdonald, 1976; Strand, 1999), l'environnement qui l'entoure (Hay & Drager, 2010), et les ses attentes et préjugés (Campbell-Kibler, 2008).

Il y a très peu d'études qui tentent de mettre en rapport ces conclusions sur la subjectivité de la

1 Inter-rater reliability

perception avec la notion d'impartialité dans les évaluations langagières. Ce fait est surprenant quand on considère que les critères des examens oraux de langue deviennent de plus en plus complexes et subjectifs (McNamara, 2001) tandis que de ces examens prennent de plus en plus d'importance en ce qui concerne la réussite scolaire et la citoyenneté (Jenkins & Parra, 2003). La plupart des études qui portent sur l'impartialité des juges se préoccupent de la fiabilité inter-juge. La supposition ici est que si plusieurs juges donnent une note très semblable à un locuteur, leur jugement est donc fiable et impartial. Toutefois, est-ce possible qu'un groupe de juges puisse avoirs les mêmes préjugés envers un certain locuteur ou groupe de locuteurs, et qu'ils puissent tous arriver à la même conclusion, même si elle n'est pas juste?

Nous avons pu trouver trois études qui ont abordé cette problématique. Toutes les trois ont démontré que les évaluations des juges ne sont pas objectives mais facilement manipulées par des éléments qui n'ont rien à voir avec les critères d'évaluation (Berthele, 2011; Rubin, 1992, Winke, Gass & Myford, 2011). Le travail présenté dans ce mémoire se voudrait être une contribution à cet axe de recherche assez mal exploité. Cette étude exploratoire examine la question suivante : les différences d'origine ethnique d'un apprenant, influencent-elles les évaluations de leur compétence langagière, que ces jugements soient implicites ou inconscients? Par ce travail nous ne cherchons pas à "démasquer" des individus et leurs jugements partiels. Nous voulons plutôt nous interroger sur les tendances au niveau sociétal qui privilégient la réussite d'un certain type de candidat au détriment d'un autre.

Comme il n'y a pas de méthodologie bien établie pour ce genre de question, nous avons développé une adaption d'un protocole déjà existant, le *Matched Guise Technique* (Lambert, 1967). Notre adaptation nous a permis de cibler l'influence de l'origine sur les évaluations des compétences langagières. Notre démarche comporte deux étapes. D'abord, nous avons demandé à de grands groupes de sujets tout-venant francophone d'évaluer trois locuteurs non-natifs, d'origine ethnique

différente, et avec un degré d'accent jugé similaire lors d'un pré-enquête. Ensuite, nous nous sommes centré sur un seul des locuteurs non-natifs et nous avons comparé les jugements des évaluateurs qui connaissaient son origine "arabe" et ceux qui ne la connaissaient pas, isolant ce facteur comme variable indépendante. Ces deux axes nous a permis d'aborder la notion d'origine dans ses dimensions "réelles" (trois locuteurs non-natifs d'origine différente) et "imaginées" (des perceptions différentes de l'origine un seul locuteur). Dans ce but, nous posons trois questions de recherche :

Dimensions "Réelles"

Nous avons proposé à des francophones tout-venant d'évaluer les compétences langagières de trois étudiantes étrangères lisant un même texte comportant les mêmes erreurs grammaticales. À partir des évaluations des juges francophones, nous nous sommes posé les questions suivantes :

1) est-ce que les juges trouveront le même *nombre total* d'erreurs et *les mêmes erreurs* chez chaque étudiante?

2.) quel est le *rapport entre les erreurs identifiées* par les juges *et les évaluations globales*[2] des compétences en langue étrangère attribuées par ces juges aux étudiantes?

Dimensions "Imaginées"

Nous nous sommes ensuite centré sur une des étudiantes étrangères, originaire de la Syrie. Nous nous sommes alors demandé si les évaluations de cette étudiante par les juges étaient différentes selon que son origine ethnique était ou non perçue comme "arabe".

3.) est-ce que les évaluations d'une même étudiante étrangère seront différentes si son origine est perçue comme "arabe"?

Le mémoire s'organise de la façon suivante. Dans un premier temps, nous présenterons les cadre

2 Les évaluations globales sont des jugements de la compétence globale de l'étudiante, notés sur une échelle de Likert de 1 à 4 (*tout à fait d'accord, d'accord, pas d'accord, pas du tout d'accord)*. Ces jugements demandent une évaluation qui prend en compte plusieurs aspects de la performance de l'étudiante pour juger : 1) leur maîtrise de la langue 2) leur intelligibilité 3) leur potentiel académique

théorique où s'inscrit notre travail. Comme cette étude se trouve au croisement de plusieurs domaines tels que la psychologie, la sociolinguistique, et la perception de la parole, nous avons dû adopter une approche pluridisciplinaire. Puis, dans un second temps, nous expliquerons les démarches méthodologiques mises en œuvre pour répondre à nos trois questions de recherche. Nous présenterons ensuite les résultats des analyses statistiques, set nous discuterons leur signification dans le contexte de l'évaluation langagière en France. Finalement, nous proposerons une conclusion où nous suggérons des orientations futures possibles.

I. APPROCHES THÉORIQUES

1. L'évaluation des compétences en langue étrangère

1.1 L'évaluation comme activité sociale

L'évaluation est une activité à laquelle chaque individu a recours au quotidien, sous des formes et des degrés différents. Réfléchir sur nos actions, décider de la meilleure ligne de conduite à suivre, former des opinions sur d'autres personnes : tout est évaluation ou mène à une évaluation.

S'il y a un lieu où l'évaluation est omniprésente, c'est l'institution éducative. On évalue ses élèves et ses étudiants dès le début de leur scolarité, et les évaluations formelles et informelles des enseignants jouent un rôle important dans le développement personnel et professionnel des élèves. On pourrait dire que l'évaluation est "le centre nerveux des pratiques d'enseignement-apprentissage" (Dervin & Salmi, 2007).

L'évaluation ne se produit pas en vase clos. Elle est, à la base, une activité sociale, liée étroitement à son contexte social, culturel, économique et politique (Sutherland, 1996). En effet, les résultats des examens représentent des compétences académiques mais peuvent aussi refléter des valeurs et des idéologies sociales (McNamara, 2001). Pendant les deux dernières décennies, les théories post-modernes ont fait ressortir le fait que, souvent, des valeurs et idéologies dominantes soutiennent des stéréotypes, soit négatifs soit positifs, envers certains groupes dans la société. On n'est pas forcement conscient de ces préjugés car ils semblent normales et légitimes (Nguyen, 1993; Gipps, 1999). Dans le cas de l'évaluation des compétences académiques, ces préjugés peuvent influencer la manière dont on juge des candidats. Ainsi, il est important de mener des études qui mesurent l'impartialité des jugements à travers tous les contextes académiques, notamment ceux qui ont des

conséquences importantes. Ce genre d'études nous permettent de prendre conscience des penchants individuels et sociétaux et de combattre la discrimination institutionnalisée (Shohomy, 2001).

1.2 L'évaluation des compétences en langue étrangère

Le sujet de l'impartialité dans les évaluations des compétences en langue étrangère est particulièrement pertinent, étant donné les intersections complexes entre langue et culture. Les notions de ce qui constitue la "compétence" en langue, ainsi que les meilleures façons de l'évaluer, sont en évolution constante, et sont influencées fortement par les valeurs dominantes de la société.

L'évaluation de la langue étrangère a beaucoup évolué depuis trente ans. L'évaluation traditionnelle visait la compétence à l'écrit, privilégiant la maîtrise de la grammaire et les structures littéraires formelles (McNamara, 2001). Pourtant, récemment, en Amérique du Nord ainsi qu'en Europe, il y a eu un changement important vers *la compétence de communication*, une notion qui met l'accent sur la capacité à communiquer de manière appropriée dans une situation de communication authentique (Canale & Swain, 1980). Cette théorie a fortement influencé les pratiques d'évaluation dans le sens d'une importance de plus en plus grande accordée aux compétences à l'oral. De plus en plus, les examens imitent des situations "authentiques" de communication, telles que des jeux de rôle, des conversations et des débats (Gipps, 1999).

Les examens de compétence en langue étrangère, reconnus nationalement, incarnent ces évolutions dans la conception de compétence. En Europe, le Diplôme d'Éducation de la Langue Française (DELF) met les candidats en position de participation à des conversations et à des débats de vingt minutes. De même, le Oral Proficiency Interview (OPI) aux États-Unis et le Canadian Language Proficiency Program au Canada, consistent à interviewer des candidats pendant trente minutes. Pour évaluer les candidats, les juges doivent prendre en compte plusieurs critères complexes tels que "peut exprimer ses idées de façon précise et nuancée" et "est capable, lorsqu'un mot lui manque, de

reformuler sa pensée de manière efficace" (Council of Europe, 2001).

Avec ce nouveau genre d'examen viennent de nouveaux défis aux processus de l'évaluation. Alors que les formes antérieures des examens consistaient souvent en questions à choix multiple ou des exercices à trou, les examens actuels sont des conversations longues et complexes. Auparavant, le travail d'un évaluateur consistait à faire des jugements objectifs du genre vrai/faux; actuellement, les juges doivent considérer toutes sortes de critères imprécis pour faire des jugements globaux de compétence (McNamara, 2001). Dans l'exemple ci-dessus, "peut exprimer ses idées de façon précise et nuancée", les juges doivent noter le degré de précision d'un candidat sur une échelle de 1 à 4. Ce type de jugement est plus complexe et holistique qu'un jugement du genre vrai/faux. De ce fait, il est difficile d'étudier l'objectivité des évaluations des candidats dans les examens de langue ayant des critères complexes (Kang, 2008).

Les conséquences de ces examens peuvent être extrêmement importantes. Par exemple, en France, le Ministère de l'intérieur et de l'immigration est chargé d'établir des critères et des processus pour qu'un immigré puisse devenir citoyen français. Dans une circulaire datée du 30 novembre 2011, le Ministère de l'intérieur a précisé que "le niveau de connaissance de la langue française requis est désormais le niveau B1 du Cadre européen commun de référence pour les langues (CECRL) du Conseil de l'Europe" (Ministère de l'intérieur, 2011, p. 3). Le candidat doit fournir la preuve de ce niveau au moyen d'un diplôme ou d'une attestation acceptable, certifiés lors d'un entretien individuel avec un agent de la préfecture.

1.3 L'impartialité dans les examens des compétences en langue étrangère

Étant donné les conséquences importantes de ces examens, il est surprenant que nous ne disposions pas d'avantage de recherches mettant en relation le fait des biais culturels ou ethniques sur les résultats de ces examens. À notre connaissance, aucune étude n'a examiné l'éventualité des biais occasionnés

par des stéréotypes culturels et linguistiques sur des examens tels que le DELF ou l'OPI. Un défi majeur de cette problématique est le grand nombre de variables en jeu. Les évaluations de la parole, lors d'une conversation, prennent en compte, parmi d'autres facteurs, l'accent, la prononciation, la fluidité, la pertinence du contenu, et la complexité. La plupart des études qui traitent de l'impartialité d'un examen mesurent la fiabilité inter-juge (entre autres, Halleck, 2008; Thompson, 2008; Surface & Dierdorff, 2003). Ces études supposent que si plusieurs juges donnent une note semblable à un locuteur, leur jugement est fiable et impartial. Toutefois, il est possible qu'un groupe de juges puissent avoir les mêmes préjugés envers un certain locuteur, et qu'ils puissent tous arriver à la même conclusion, même si elle n'est pas juste.

L'éventualité d'un biais, social ou ethnique, partagé par plusieurs évaluateurs, est un sujet extrêmement délicat. Toutefois, vu ces conséquences sociales et scolaires, cette question mérite d'être examinée. Dans la section suivante, nous allons examiner les études qui explorent les influences diverses jouant un rôle dans l'évaluation afin d'aider notre réflexion.

2. Apports des disciplines différentes à la question de l'impartialité dans l'évaluation

Le sujet traité dans ce travail est celui de la discrimination institutionnalisée[3] et son influence éventuel sur les résultats des examens de compétence langagière. Comme il n'y a pas de corpus théorique et empirique qui aborde cette question, nous devons rechercher des théories provenant de plusieurs disciplines pour définir notre cadre théorique.

Chaque domaine traitant de sujet des évaluations et des prises de décisions adopte une approche différente, fondée sur de suppositions de base spécifiques et de méthodologies particulières. Dans la section suivante, nous allons considérer plusieurs approches différentes afin d'établir notre cadre théorique. Nous allons partir du principe, généralement reconnu, que les stéréotypes des groupes sociaux, dont les locuteurs sont membres (ou sont perçus comme membres), peuvent exercer une influence sur la façon dont ils sont perçus (entre autres, Beebe, 1981; Thakerar & Giles, 1981; Williams, 1976). Nous allons cependant nuancer cette notion en nous posant les deux questions suivantes :

1) À quel point les juges sont-ils conscients de ces influences?

2) Est-ce que les stéréotypes sociaux affectent la *perception* elle-même des personnes? Ou bien est-ce qu'ils impliquent des *traitements* différents d'une perception qui reste inchangée?

Les réponses à ces questions aideront à définir notre choix de méthodologie pour cette étude exploratoire.

2.1 Le modèle à deux voies

La question du poids relatif des aspects conscients et inconscients dans nos processus de pensée, se pose dans plusieurs domaines de recherche, y compris la psychologie, les neurosciences et la cognition

3 Par 'institutionnalisé', nous ne voulons pas dire une discrimination explicite par un gouvernement. Nous employons ce terme dans un sens plus large. Ce sens comprend des pratiques de discrimination implicite, par exemple le fait d'embaucher souvent un certain type de candidat pour un poste, au détriment d'un autre type de candidat ayant des qualifications similaires (Jones, 2000)

sociale. Une des questions centrales concerne le contrôle que peuvent avoir les être humaines sur leurs propres pensées et les prises de décision. Quelles sont les forces qui influencent notre perception, notre réflexion et notre action? À quel point nos décisions sont-elles le résultat d'un processus conscient et à quel point sont-elles automatiques ou inconscientes?

Par le passé, la doctrine du rationalisme dominait la pensée psychologique. Ce principe affirme que le jugement est le résultat d'un raisonnement conscient et d'une réflexion volontaire (Gauthier, 2011). En d'autres termes, l'esprit humain aurait une faculté qui lui permettrait de fixer des critères de vérité et d'erreur, de discerner le bien et le mal et de diriger ses jugements par la volonté. Pourtant, au siècle dernier, certains psychologues tels que Freud et Skinner ont défié cette doctrine, affirmant que nous sommes principalement dominés par des forces inconscientes ou associatifs. Dans la théorie de l'inconscient de Freud, par exemple, nos actions sont fortement influencées par des motivations inconscients et nos réflexions dites 'conscientes' proviennent toujours de l'inconscient. Selon cette vision, le comportement et la pensée consciente de chaque individu sont fortement influencés par des forces dont il ou elle ne se rend pas compte.

La psychologie contemporaine a tendance à trouver un équilibre entre ces deux pôles. Un grande nombre de psychologues adhèrent aujourd'hui à un *modèle des processus à deux voies*[4] (Campbell-Kibler, 2010), qui rejette la notion que nous sommes dominés entièrement par le conscient ou par l'inconscient. Nos décisions seraient plutôt le résultat d'interactions entre des processus implicites et explicites. Le débat aujourd'hui se cristallise autour de la question suivante : dans une situation donnée, quel serait le poids des processus conscients et des processus inconscients? Dans la vie quotidienne, sommes-nous dominés davantage plus par des forces conscientes ou inconscientes? (Bargh & Williams, 2006).

4 dual process model

Depuis vingt ans, des avancées technologiques nous ont permis de mieux comprendre ces interactions. La neuropsychologie utilise des méthodes d'exploration de l'activité cérébrale pour mieux comprendre le fonctionnement des structures neurologiques impliquées lors de la prise de décision et du jugement. Il s'avère que les données d'un grand nombre d'études soutiennent la notion d'un modèle des processus à deux voies, démontrant que plusieurs réseaux neurologiques, conscients et inconscients, rationnels et émotionnels, communiquent entre eux dans une prise de décision (entre autres, Gazzaniga, M., 2005; de Damasio, 2007). Ainsi, nos décisions sont influencées par plusieurs facteurs dont on ne se rend forcement pas compte.

2.2 L'automaticité

Des chercheurs dans le domaine des sciences cognitives ont commencé à approfondir les processus qui sous-tendent cette notion d'inconscient. Les théories de *l'automaticité* soutiennent que pour faire face au bombardement des informations du monde externe, le cerveau apprend à faire certaines tâches automatiquement (Bargh et Chartrand,1999; Bargh et Williams, 2006; Von Hippel, Sekaquaptewa & Vargas, 1995). Au fil du temps, des réponses à certains stimuli répétés deviennent automatisées, ce qui nous permet de fonctionner de façon plus efficace (Shiffrin, Dumais & Schneider, 1981). Par analogie, apprendre à marcher prend beaucoup d'énergie de de concentration. Par contre, quelques années plus tard, on n'y pense plus parce que nos muscles ont appris à s'activer automatiquement. De même, les études sur l'automaticité suggèrent que certains processus de pensée sont réalisés sans attention parce qu'ils sont devenus force d'habitude.

Une méthode, développée par l'Université de Harvard, pour mieux comprendre ces processus automatiques, s'appelle le "Implicit Association Test" (IAT). L'IAT se présente comme "une tâche de catégorisation informatisée censée appréhender de façon indirecte les forces d'associations entre concepts" (Chassard, 2006 : 15). Le principe sur lequel repose l'IAT est le suivant : il devrait être

d'autant plus facile (*e.g.* plus rapide) d'associer deux concepts que ceux-ci sont plus fortement associés dans l'esprit. Dans un IAT, le sujet doit classer – le plus rapidement possible en faisant le moins d'erreurs possible – des exemplaires représentant deux catégories. Dans l'exemple INSECTE-FLEUR, on présente aux participants deux réponses possibles (BON et MAUVAIS) et ils doivent associer des exemplaires de la catégorie INSECTES et de la catégorie FLEURS à ces deux réponses BON ou MAUVAIS le plus rapidement possible. Plus précisément, le sujet voit une série des mots comme CAFARD, PAPILLON et ROSE sur l'écran et il doit cliquer sur BON ou MAUVAIS le plus rapidement possible. Le logiciel mesure le temps nécessaire pour faire chaque association, calculant le poids du lien dans le cerveau. Le lien fait le plus rapidement démontrent une forte association dans le cerveau. Dans une étude de Greenwald, McGhee et Schwartz (1998), les sujets étaient généralement plus rapides à mettre ensemble FLEURS et BON (INSECTES et MAUVAIS) qu'à mettre ensemble INSECTES et BON (FLEURS et MAUVAIS). Les chercheurs soutiennent que ces résultats démontrent que les sujets préfèrent généralement les fleurs aux insectes.

Plusieurs chercheurs ont utilisé les IAT pour mesurer la force d'association entre les concepts sociaux. Par exemple, un IAT mesure la force d'association entre BON et MAUVAIS, JAPONAIS et COREENS et une autre vise l'association entre BON et MAUVAIS, BLANC-AMÉRICAIN et NOIR-AMÉRICAIN. Les résultats des études utilisant ces tests mettent en évidence la présence d'une forte association mentale qui rend plus rapide les réponses dans une certaine direction : par exemple, les Japonais qui font le test JAPONAIS-COREENS associent plus rapidement JAPONAIS-BON que COREEN-BON; les Coréens qui font le même test associent plus rapidement COREEN-BON que JAPONAIS-BON (Greenwald et. al, 1998).

Les résultats du test IAT ne sont pas toujours corrélés à des réponses explicites des participants. Certains chercheurs avancent l'hypothèse qu'il existe un "filtre" empêchant les participants d'exprimer

leurs vraies pensées à des questionnaires comportant des questions délicates sur les stéréotypes raciaux (Greenwald et al, 1998). Pourtant, l'IAT est souvent capable de prédire le comportement des individus. McConnell et Leibold (2001) ont démontré ce phénomène dans une étude qui explorait le rapport entre des réponses à un "Race IAT" (mesurant les associations entre les concepts BON et MAUVAIS et les personnes BLANCHES et NOIRES) et les actions des participants envers des expérimentateurs noirs et blancs. Les participants ne savaient pas que leurs interactions avec ces expérimentateurs avaient été filmées et ensuite codées par des psychologues. Les résultats démontrent une forte corrélation entre les résultats du IAT et le comportement des participants envers les expérimentateurs noirs et blancs.

Les critiques de l'IAT souligne que les résultats de ce test ne représentent pas forcement des préjugés individuels. En effet, ils pourraient simplement refléter des stéréotypes vus souvent dans la société (Fazio & Olson, 2003). Par exemple, si l'on voit souvent des représentations négatives envers des personnes noires dans les médias, cette association va ressortir automatiquement lors d'un IAT, même si elle ne correspond pas à des croyances individuels. Pourtant, comme ces associations automatiques peuvent influencer notre comportement, elles méritent d'être considérées dans des études de jugement. Ainsi, pour mieux comprendre les influences des stéréotypes sur les évaluations langagières, il semble important de trouver un protocole qui prend en compte les associations automatiques et les penchants inconscients des juges.

2.3 Perception de la parole

Si nous voulons développer un protocole qui prend en compte ces associations automatiques, il est important de préciser à quel moment interviennent nos préjugés : au moment de l'encodage d'un stimuli ou plus tard, dans son traitement? Cette question est abordée par plusieurs disciplines, y compris la perception de la parole.

Le domaine de la perception de la parole s'occupe depuis longtemps des processus par lesquels les

humains sont capables d'interpréter et de comprendre les sons utilisés dans le langage. Cette étude est reliée aux champs de la phonétique, de la phonologie en de la linguistique, de de la psychologie cognitive. Pourtant, traditionnellement, l'accent a été mis sur les mécanismes mis en œuvre dans la production de la parole avec la perception mise en second plan (Thomas, 2002). En se concentrant sur la nature articulatoire, on privilégiait le point de vue du locuteur sur celui de l'auditeur. En outre, quand l'audition était traitée, ce n'était pas en relation avec les aspects sociolinguistiques mais plutôt dans une perspective strictement cognitive: quels sont les mécanismes mentaux responsables dans le traitement du son dans la parole?

Les évolutions récentes des domaines de la psychologie, de la sociolinguistique et de la cognition sociolinguistique ont beaucoup apporté à ce domaine. L'attention mise sur le rôle des processus inconscients soulèvent des questions de subjectivité dans la perception. Effectivement, de plus en plus d'études s'occupent du côté perception, identifiant les influences non-auditives sur l'ouïe.

Ces études se regroupent autour de trois thèmes principaux :

1) l'aspect multimodale de l'audition

2) l'importance de l'environnement et du contexte

3) l'influence des représentations internes

2.3.1 l'aspect multimodale de l'audition

La perception de la parole ne comprend pas seulement les structures auditives. En effet, notre vision joue aussi un rôle important dans notre interprétation du bruit. *L'effet McGurk* (1976) est un phénomène reconnu pour avoir montré une interférence entre l'audition et la vision lors de la parole. En effet, quand on présente une vidéo montrant une personne prononçant le phonème (/ga/) alors que la bande sonore diffuse l'enregistrement d'un autre phonème (/ba/), les auditeurs ont l'impression d'entendre un troisième phonème intermédiaire (/da/).

Ce résultat a mobilisé d'autres chercheurs pour appliquer ce phénomène analogue à d'autres influences multimodales à partir de stimuli évocateurs de catégories sociales. Par exemple, dans une étude récente, Hay, Warren et Drager (2006) faisaient écouter un même enregistrement à différents sujets. Chaque groupe voyait une photo différente du locuteur, représentant une variété d'âges et de classes sociales. Les chercheurs constatent que la photo influence fortement les perceptions des phonèmes. Ils concluent que "le fait que les aspects sociaux des photos ont une influence sur la performance des évaluateurs suggère fortement que les évaluateurs sont sensibles aux informations sociales lors de la perception de la parole.[5]" (p. 480, notre traduction). Ces conclusions soutiennent les résultats d'autres études similaires, qui démontrent que des caractéristiques sociolinguistiques perçus d'un locuteur ont un impact sur la façon dont des auditeurs perçoivent ces paroles (Strand, 1999).

Il nous a paru important ici d'apporter un éclairage sur l'innovation que représente cette dernière méthodologie. En effet, elle présente certaines analogies avec à un protocole classique, qui s'appelle le *Matched Guise Technique* (MGT) de Lambert (1967). Dans l'étude originale, un locuteur multilingue est enregistré, lisant à haute voix un même texte traduit en français et en anglais. Ensuite, des évaluateurs (canadiens anglophones et canadiens francophones) notent leurs attitudes envers certains caractéristiques du locuteur, tels que sa personnalité ou son statut social. L'objectif réel est dissimulé parce que les évaluateurs pensaient qu'ils comparent deux locuteurs différents mais, en réalité, il n'y en a qu'un seul. Lambert trouve que les évaluateurs anglophones jugent le locuteur moins favorablement lorsqu'il parle en français. Inversement, les évaluateurs francophones le jugent moins favorablement lorsqu'il parle en anglais.

L'étude de Hay, Warren et Drager (2006) modifie ce protocole dans la mesure où ils utilisent un même enregistrement avec plusieurs groupes de juges. En outre, le but de ces auteurs n'est pas de

5 the fact that social aspects of the photos affect participants' performance provides strong evidence that individuals are sensitive to social information in speech perception

recueillir l'influence du langage perçu sur la formation de l'image sociale qu'on attribue au locuteur, comme c'est le cas pour le Matched Guise. Ils explorent plutôt l'influence inverse, c'est-à-dire l'influence de l'image sociale sur la perception de la parole elle-même. Ce n'est donc pas l'enregistrement qui distingue les groupes mais le fait que les expérimentateurs donnent à chaque groupe une photo différente pour recréer l'identité du locuteur. De ce fait, dans le protocole de Lambert, chaque groupe de juges écoutent un stimulus différent, alors que dans le protocole de Hay et Drager (2006), chaque groupe de juges écoutent le même stimulus.

Cette évolution représente un changement majeur de perspective concernant la perception. Avec le protocole original du MGT, on prend pour acquis que l'on entend la même chose et que les variétés de parole vont activer des représentations internes, mobilisant parfois des préjugés. Avec la nouvelle adaptation, on met l'accent sur la subjectivité de l'encodage des informations, démontrant que ce n'est pas seulement le signal auditif que l'on traite pour comprendre la parole mais aussi les perceptions visuelles.

2.3.2 l'importance de l'environnement et du contexte

À la suite des études montrant l'importance des informations visuelles et auditives, plusieurs chercheurs ont tenté de savoir si des représentations internes pouvaient être activées de façon moins explicite. Une étude en particulier souligne l'influence implicite du contexte. Hay et Drager (2010) ont examiné la perception des sons des voyelles en australien et en néo-zélandais. Les évaluateurs devaient écouter un enregistrement d'un homme et associer son accent avec un des six autres enregistrements : six voix représentant les variants d'accents néo-zélandais et australiens. Dans cette expérience, les expérimentateurs ne donnaient aux évaluateurs aucune information sur l'identité, l'apparence physique, ou l'origine du locuteur. Ils se contentaient de faire semblant de découvrir un animal en peluche dans un tiroir ouvert pour les besoins de l'expérience et de le plaçait sur la table. Cet animal était un kiwi,

symbolique de la Nouvelle-Zélande, ou bien un koala ou un kangourou, symboliques de l'Australie. Les résultats démontrent que le type d'animal exerce une influence notable sur les réponses des évaluateurs. Ainsi, les évaluateurs ayant perçu un kiwi ont plus tendance à entendre un accent néo-zélandais, alors que ceux qui ont perçu un koala ou un kangourou ont tendance à entendre un accent australien.

Ces résultats corroborent les conclusions de l'étude de Hay, Warren et Drager. (2006). En effet, lors de leur analyse des résultats, ils avaient remarqué que les évaluateurs interagissant avec un expérimentateur américain distinguaient mieux les phonèmes américains que ceux qui interagissaient avec un expérimentateur néo-zélandais. Ils en ont conclu que l'expérimentateur lui-même devait être considéré comme une variable indépendante dans les études de la perception de la parole.

2.3.3 L'influence des représentations internes

Ces dernières études suggèrent que tous les éléments de l'environnement qui nous entourent peuvent jouer un rôle dans la perception de la parole. Pourtant, ce n'est pas toujours notre système visuel qui nous influence. En effet, nos représentations internes peuvent également jouer un rôle.

Niedzielski (1999) a utilisé une variante du MGT pour examiner les représentations internes des évaluateurs sur des accents canadiens et américains. Elle demandait à des évaluateurs d'associer une voix entendue sur un enregistrement avec une des voyelles entendues sur des courts clips-audio. Ces voyelles représentaient des voyelles typiques des locuteurs de Detroit et des voyelles typiques des locuteurs de Toronto, une ville juste au nord de Detroit, au Canada. Utilisant un même enregistrement, elle expliquait à la moitié des évaluateurs que le locuteur était du Canada et à l'autre moitié qu'il était de Detroit. Elle constate que les évaluateurs entendent l'accent correspondant à l'origine présumée, et imaginaire, indiquée par l'expérimentateur. En effet, les évaluateurs pensant écouter un canadien identifiaient les phonèmes typiquement canadiens; les évaluateurs pensant écouter un américain

identifiaient les phonèmes typiquement américains. Pourtant, tous les évaluateurs écoutaient exactement la même chose. Ces résultats montrent que le placement de la frontière entre les phonèmes d'une langue dépende des représentations sociales internes des interlocuteurs.

Campbell-Kibler (2008) a employé une méthode similaire pour examiner la perception des variations du phonème 'ing' aux États-Unis. Ses résultats corroborent ceux de Niedzielski (1999). Elle trouve que les auditeurs déplacent leurs perceptions des variantes selon l'identité fournie du locuteur. Elle avance la notion du "listener agency" (le pouvoir de l'auditeur), estimant que les attentes d'un auditeur influencent ce qu'il perçoit et donc la construction du sens dans les interactions. Effectivement, les auditeurs incorporent leur image du locuteur dans leur compréhension de la parole.

Dans le domaine de la parole, il existe donc une accumulation croissante d'indices que les attentes de l'auditeur influencent la perception. "Les stéréotypes des groupes sociaux dont les locuteurs sont membres (ou sont perçus comme membres), ont une influence sur la façon dont les variantes de leur parole sont perçues[6] (Niedzielski, 1999 : p. 62).

2.4 Liens avec l'impartialité des jugements sur la compétence langagière

Jusqu'à présent, le rapport entre les influences inconscientes sur la perception de la parole et l'évaluation de la compétence en langue étrangère est resté relativement inexploré. Comme nous l'avons évoqué en section 1.2, il y a très peu d'études qui mettent en relation les conclusions des études en perception de la parole avec des questions sur l'impartialité des examens de compétence langagière. Les études qui établissent une fiabilité inter-juge éliminent effectivement la subjectivité individuelle mais ne touchent pas à la subjectivité collective, résultant des possibles effets des préjugés institutionnalisés et partagés au niveau sociétal. Nous n'avons pu trouver que trois études qui se

6"stereotypes about the social groups speakers are members of (or are believed to be members of) have an influence on how their language varieties are perceived

trouvent au croisement de ces deux domaines et leurs conclusions affirment le besoin de mieux développer cette piste de recherche.

Dans la première, Rubin (1992) examine l'impact des stéréotypes asiatiques sur les évaluations des enseignants de l'université.[7] Rubin a enregistré un discours de quatre minutes d'un enseignant américain (blanc, anglophone natif) et l'a fait écouter auprès de 62 étudiants. Les étudiants écoutaient la bande sonore en même temps qu'ils voyaient une photo de l'enseignant. Une moitié des étudiants voyait l'image d'une personne blanche et l'autre moitié voyait une photo d'une personne asiatique. Ensuite, tous les étudiants remplissaient un questionnaire qui leur demandait de noter la force de l'accent ainsi que leur compréhension du contenu du discours. D'une manière surprenante, Rubin trouve que bien que l'enseignant ait un accent américain, les étudiants qui voyaient un visage asiatique percevaient un accent étranger. De façon inattendue également, l'auteur observe une corrélation entre la perception d'un accent étranger et la compréhension du discours enregistré – les étudiants percevant un accent étranger obtenaient des notes plus basses sur les questions de compréhension du discours. Ces résultats suggèrent que non seulement les étudiants entendent différemment la même voix mais aussi ils la comprennent différemment. En bref, lorsqu'ils voient un visage asiatique associé à l'extrait du discours, les étudiants entendent un accent asiatique et comprennent moins bien le contenu du discours.

Dans la deuxième étude, Berthele (2011) mesure l'influence de l'origine ethnique et de l'*alternance codique*[8] sur les évaluations des professeurs dans une école primaire en Suisse. Le chercheur a recueilli un corpus de trois textes authentiques en français à l'oral et en utilisant un logiciel, il a ré-enregistré chaque texte en y incluant des mots allemand, créant ainsi un phénomène d'alternance codique. Chacun des 6 textes (3 enregistrements non modifiés et 3 avec l'alternance codique ajoutée) ont été

7 Les enseignants étaient des "teaching assistants" – des assistants à un professeur titurlaire.

8 Un alternance de deux ou plusieurs codes linguistiques (langues, dialectes, ou registres) dans un seul énoncé

attribuées à deux profils d'un locuteur : soit un élève suisse-allemand natif soit un élève serbe ayant grandi en suisse. Ensuite, 157 professeurs d'une école élémentaire ont écouté les enregistrements et leur ont attribué des notes pour les critères suivants : la fluidité, la précision et le potentiel académique de l'élève. Les résultats démontrent que lorsque les professeurs pensent que le texte provient d'un élève serbe, ils donnent une meilleure note, sauf dans le cas de l'alternance codique. Dans ce cas, ils sont plus sévères. L'inverse apparaît pour les textes perçus comme provenant d'un élève suisse-allemand : en cas d'alternance codique, les professeurs se révèlent moins sévères. Cette étude démontre que les jugements d'acceptabilité sur certaines erreurs, telles que l'utilisation de mots d'une autre langue, peuvent changer selon l'identité perçue du locuteur. Ces conclusions peuvent avoir des conséquences importantes pour les études sur l'évaluation des compétences langagières parce qu'elles soulignent la subjectivité des jugements des évaluateurs, mêmes des enseignants formés.

Dans la troisième étude, Winke, Gass et Myford (2011) ont examiné l'impact que pourrait avoir la connaissance d'une langue étrangère sur les évaluations des locuteurs natifs de cette langue. Cent sept juges anglophones, qui tous ont suivi des cours soit de chinoise soit de coréen soit d'espagnol, ont écouté soixante-douze locuteurs natifs, chinois, coréens ou espagnols qui parlaient l'anglais. Les juges ont attribué des notes à chaque locuteur pour leur compétence en anglais. Après avoir analysé ces évaluations, les chercheurs ont constaté que les juges avaient tendance à mieux évaluer les locuteurs ayant une langue maternelle que le juge lui-même connaissait.

Avec ces trois dernières études, nous constatons qu'il est possible d'isoler certaines variables pour explorer leurs influences sur des jugements. Rubin (1992) et Berthele (2011) ont isolé la variable *identité perçu* du locuteur en fournissant des identités différentes d'un même locuteur. Winke et. al (2011) ont isolé la variable *connaissance de la langue maternelle du locuteur* en examinant le rapport entre les langues connues par des juges et la langue maternelle du locuteur.

Étant donné le petit nombre d'études similaires, il est très difficile de généraliser ces résultats. En outre, il n'existe pas de protocole bien établi et il y a donc très peu d'homogénéité entre les méthodes. Le nombre de variables possibles semble infini : l'environnement, les expériences antérieures des juges, l'information donnée sur l'origine du locuteur, le type du discours (formel ou non-formel), la longueur du clip, etc. Si on ajoute à cela l'évolution constante de la technologie, alors on comprend que l'exploration de ce champ nouveau n'en est qu'à ses toutes premières tentatives et que l'essentiel du travail reste à faire.

En dépit de ces défis, nous pensons que les possibilités pour mieux connaître le processus d'évaluation en langue étrangère sont immenses et que leur exploration répondrait à des obligations déontologiques en matière d'équité en notation. Cette exploration devrait être facilitée par l'élaboration d'un cadre théorique partagée, point qui sera développé dans les lignes suivantes.

3. Cadre Théorique

3.1 La cognition sociolinguistique

Nous arrivons maintenant à une clarification explicite de notre cadre théorique. Comme nous l'avons évoqué en 1.2, il n'existe pas de corpus de recherche bien établi qui touche à la fois à la perception de la parole et l'évaluation des compétences langagières. Pourtant, les nouvelles théories à propos de l'automaticité dans le domaine des sciences cognitives ont ouvert de nouvelles dimensions en sociolinguistique, ce qui engendre des questions importantes sur l'impartialité de nos jugements. Nous avons donc choisi d'ancrer notre étude dans les théories de la cognition sociolinguistique. Ce domaine, relativement nouveau et influencé par des avancées en sciences cognitives et en sociolinguistique, explore, entre autres choses, les mécanismes individuels qui nous permettent d'encoder et de traiter de la variation sociolinguistique (Geeraerts, Kristiansen, & Peirsman, 2010; Campbell-Kibler, 2010). Ces mécanismes peuvent être automatiques ou inconscients et sont influencés par des traits caractéristiques (réels ou imaginés) du locuteur.

Selon Hippel, Sekaquaptewa & Vargas (1995), ces mécanismes se produisent au moment de l'encodage des informations. Ils estiment qu'il existe un type *d'interférence automatique* entre la présence d'un stimulus et l'encodage de ce stimulus. En effet, nos attentes, y compris nos stéréotypes, façonnent notre manière de voir et d'entendre, influençant la quantité et la nature de l'information encodée. Un stéréotype très fort peut inhiber ou promouvoir certains éléments du stimuli, rendant l'évaluation de ce stimuli forcément subjective.

Bargh et Williams (2006) soulignent que ces processus dits "subjectifs" ne sont pas nécessairement négatifs. En effet, ils sont très efficaces et nous permettent de fonctionner et de prendre des décisions dans un monde où on est bombardé d'un flux constant d'information. Dans ce cas, notre système d'encodage automatique est fortement suscité pour trier des informations pertinentes.

Toutefois, il faut aussi considérer les conséquences plus négatives de ce type de traitement. Comme ces processus se produisent automatiquement, il est très difficile de se rendre compte des associations automatiques qu'ils impliquent. Les préjugés peuvent être construits au niveau individuel mais sont le plus souvent influencés par les courants dominants de la société. De ce fait, même si nous ne partageons pas certains préjugés à titre individuel ou de façon explicite, ces associations dominantes qui circulent dans la société peuvent biaiser nos évaluations.

3.2 Les Concepts Clés

Utiliser ce cadre théorique exige une reformulation des concepts clés tels que "préjugés" et "stéréotypes". En effet, l'accent mis sur les processus automatiques et la subjectivité de l'encodage des information changent fondamentalement leurs sens. Afin d'éclaircir ce changement de perspective, nous avons comparé les définitions plus traditionnelles établies par les théories de la psychologie avec les définitions proposées par des chercheurs en cognition sociolinguistique (tableau 1). Évidemment, ce sont ces dernières définitions qui seront utiles à l'étude présentée dans ses lignes.

Tableau 1 : Comparaison des définitions des concepts clés.

Concept	**Définition Traditionnelle**	**Définition suivie dans cette étude**
Communication	Le fait de transmettre quelque chose. *"an act or instance of transmitting" (Merriam-Webster Dictionary)*	Un processus réciproque où le locuteur et l'auditeur jour un rôle dans la construction du sens. *A reciprocal process where both speaker and listener play an important rôle (Rubin, 1992)*
Perception	Des processus conscients et systématiques pour interpréter le monde et planifier des actions. *"people are consciously and systematically processing incoming information in order to construe and interpret their world and plan and engage in courses of action" (E.J. Langer, 1978)*	Des processus pour la plupart automatiques ou inconscients. *"most of moment-to-moment psychological life must occur through nonconscious means if it is to occur at all" (Bargh et Chartrand, 1999)*
Stéréotypes	Un ensemble de généralisations sur un groupe de personnes ou une catégorie sociale. *"a set of generalisations about a group of people or a social category" (Psychology Dictionary)*	Des associations automatiques, apprises par habitude et activées lors d'une rencontre. *"stereoypes are overlearned and are automatically activated upon encounters" (von Hippel, Silver, & Lynch, 2000)*
Préjugés	Des attitudes négatives envers des membres d'exogroupes. *"negative attitudes toward specific outgroups" (Allport, 1954)*	Des interférences automatiques au moment de l'encodage plutôt que des croyances conscients *"encoding processes rather than the evaluative content of a perceiver's beliefs about an out-group" (von Hippel, 1995)*

II. PRÉSENTATION DES PROTOCOLES EXPÉRIMENTAUX

Les protocoles de cette étude s'inscrivent dans la tradition des protocoles type "Matched Guise Technique" (MGT), décrite en section 2.3.1 (Approches Théoriques). Afin d'examiner l'influence des stéréotypes dans les évaluations langagières, nous avons cherché un protocole nous permettant de contrôler le plus de variables possibles. Ainsi, nous avons développe une adaptation du protocole MGT qui nous permettraient de cibler les variables *d'origine ethnique*[9], tout en préservant le plus de similarité possible entre les locuteurs.

Afin d'éclaircir la préparation et le déroulement de ce protocole, nous commencerons par décrire le contexte de l'étude dans son espace linguistique et culturel. Dans un deuxième temps, nous détaillerons la préparation de l'expérience :

1. la présélection de trois étudiantes étrangères ayant des origines ethniques différentes
2. la création du texte destiné à être enregistré par les trois étudiantes étrangères et à être évalué par des juges francophones
3. les conditions d'enregistrement des trois bandes sonores.

À la suite de la description de la préparation de l'expérience, nous présenterons le déroulement de l'expérience, où les juges francophones écoutent un des trois enregistrements du texte et évaluent la compétence en français de l'étudiante étrangère entendue. Ensuite, nous définirons six conditions expérimentales différentes. Enfin, nous expliciterons la saisie et le traitement des données.

9 Ce terme est utilisé dans ce mémoire pour dénoter un ensemble de représentations sociolinguistique d'un individu. Ces représentations peuvent comprendre l'origine nationale, culturelle ou linguistique de l'individu, et peuvent être fondées sur les faits réels et imaginés. Elles comprennent souvent des stéréotypes.

1. Contexte de l'étude et son impact sur les hypothèses de l'étude

1.1 Le contexte linguistique et culturel en France

Avant d'exposer la méthodologie de notre étude, il est nécessaire d'en présenter le contexte de l'étude dans son espace linguistique et culturel. La France est un pays caractérisé par un brassage de langues et de cultures. Cette diversité est une source de richesse pour le pays mais elle crée parfois des tensions. La nature exacte de ces tensions varie considérablement selon le lieu, l'époque et la sous-communauté considérée. Pourtant, malgré cette importante variabilité, des tendances générales peuvent être constatées.

Il y a eu, dans l'histoire de France, plusieurs vagues d'immigration d'un groupe particulier, qui ont provoqué des sentiments de peur ou de méfiance de la part des citoyens français. En 1895, après une forte immigration des italiens et des allemands en France, Le Bon remarquerait que "l'immigration de ces voisins est fatale...si ces invasions ne s'arrêtent pas, il faudra un temps bien court pour qu'en Franc un tiers de la population soit devenu allemand et un tiers italien" (Le Bon, 1895 : p. 124). Cinquante ans plus tard, c'est une immigration polonaise qui a provoqué ce genre de sentiment : "c'est une véritable invasion méthodique...des cités (qui deviennent) de véritables villages étrangers, où le français n'est pas compris" (Ariès, 1948 : p. 110-111). Tout en n'étant pas partagées par tous les français, ces appréciations que nous rapportons soulignent comment les différentes vagues d'immigration peuvent solliciter des sentiments de peur et de méfiance et aboutir à des contextes de discrimination au niveau sociétal.

Actuellement en France, une part importante de la population immigrée vient des pays maghrébins (Tunisie, Maroc, Algérie, Mauritanie et Libye). Selon plusieurs études, il existe une discrimination importante envers ces immigrés et leurs enfants nés sur le sol français. À niveau socioprofessionnel égal, les personnes maghrébines ont plus de difficulté à trouver du travail (Silberman, Alba et Founier,

2006). Selon Simon et Clément (2006), "à diplôme et origine sociale comparables, les descendants d'immigrés maghrébins ont moins de chance de trouver un emploi que les jeunes d'origine portugaise ou française" (p. 1).

Dans cette étude, nous visons les influences possibles des associations négatives de certains accents ou origines dans un contexte d'évaluation langagière. Dans ce but, il est important de prendre en compte le paysage linguistique et culturel unique de la France. Notre recherche suggère qu'actuellement en France, il existe une discrimination institutionnalisée envers la population maghrébine. Ces sentiments peuvent exister autant au niveau individuel que sociétal.

1.2 Classement des groupes linguistiques et culturels en France

Nous cherchons à isoler l'origine perçue comme variable dans notre étude. Il nous fallait donc choisir une terminologie pour identifier l'origine des trois étudiantes étrangères qui ont lu le texte, qui devait être évalué par les juges. Évidemment, classer ces étudiantes selon des critères linguistiques et culturels est une activité à connotation politique. Elle est pourtant nécessaire dans le contexte de notre étude puisque nous cherchons à déterminer quels stéréotypes bien connus de la société française pourraient avoir une influence sur l'évaluation des compétences en langue étrangère. Nous sommes pleinement consciente du risque qu'il y a à participer à la perpétuation des stéréotypes du simple fait de les nommer, même avec distance.

Certains chercheurs suggèrent qu'il existe un système des nomenclatures qui, en France, distinguent des individus selon leur langue maternelle, leur couleur de peau, et l'origine de leur famille (Omi & Winant, 1994; Simon, 2008). La formation de ce système de catégories est historique, politique, et économique et la catégorisation des populations est donc "en prise avec le monde social, enregistrant ses évolutions" (Simon, 2008, p. 541). Selon Simon et Clément (2006), les quatre nomenclatures raciales dominant actuellement dans les pratiques linguistiques ordinaires en France

sont : "Blanc", "Noir", "Arabe" et "Asiatique". Nous avons donc contacté des étudiants qui pourraient "représenter" trois de ces dernières catégories : "Blanc", "Arabe" et "Asiatique".

1.3 Formulation d'une hypothèse située dans son contexte spécifique

Nous pensons qu'il est important de formuler une hypothèse qui soit ancrée dans son temps et son espace linguistique et culturelle. Les tendances générales, mises en évidence dans les sections précédentes, nous permettent d'établir les hypothèse suivantes :

Étape 1 : Trois étudiantes étrangères ayant des origines ethniques différentes lisant un même texte

Question 1 : Étant donné trois étudiantes étrangères, correspondant aux catégories "blanc", "arabe", et "asiatique", lisant un même texte comportant les mêmes erreurs grammaticales, nous nous demandons si les juges trouveront le même nombre total d'erreurs et les mêmes erreurs chez chaque étudiante.

Hypothèse 1 : Vu les préjugés défavorables qui circulent dans la société française vis à vis de l'immigration arabophone, nous faisons l'hypothèse que les juges trouveront davantage d'erreurs chez l'étudiante d'origine syrienne.

Question 2 : Nous nous demanderons également quel est le rapport entre les erreurs identifiées et les évaluations globales des compétences en français langue étrangère des étudiantes (maîtrise de la langue, prononciation, et potentiel académique)?

Hypothèse 2 : Nous supposons que ces évaluations globales seront d'autant meilleures que les erreurs identifiées seront moins nombreuses.

Étape 2 : Une étudiante étrangère lisant un texte. Comparaison des juges qui pensent qu'elle est "arabe" et ceux qui pense qu'elle n'est pas "arabe"

Nous avons divisé les juges écoutant un même enregistrement en deux catégories :

a) ceux qui perçoivent une origine "arabe" de l'étudiante étrangère

b) ceux qui perçoivent une origine "non-arabe" de ce même étudiante étrangère

Question 3 : Nous nous demandons si les évaluations d'une même étudiante étrangère sont différentes si son origine ethnique est perçue ou donné comme "arabe"?

Hypothèse 3: Nous supposons que dans le cas de l'étudiante perçu comme "arabe", les évaluations seront plus sévères si les juges perçoivent cette origine.

2. Préparation de l'expérience

2.1 Le protocole de présélection des étudiantes étrangères à enregistrer les énoncées

Afin d'isoler l'accent comme variable, nous avons cherché trois étudiantes étrangères représentant "Blanc", "Asiatique" et "Arabe", avec des accents différents typiques de leur communauté mais très semblables sur d'autres traits. En effet nous devions contrôler l'âge, le genre et le niveau de français et il nous semblait important de nous assurer aussi que toutes les étudiantes avaient un niveau semblable d'intelligibilité. Si une des trois avait une prononciation moins intelligible que les autres, cette différence pouvait rendre ses énoncées plus difficiles à comprendre et donc influencer les réponses des juges.

La distinction entre *accent* et *prononciation* n'est pas toujours claire. Où fixons-nous la limite entre un accent marqué et une prononciation inexacte? La meilleure solution à cette question très complexe a été de suivre un processus rigoureux de présélection des étudiantes étrangères, fondé sur le critère de l'*intelligibilité,* c'est à dire "la qualité de ce qui peut être compris, saisi aisément" (Larousse, 2013).

Notre protocole a suivi deux étapes. Dans un premier temps, sept étudiantes ont enregistré un texte à juger et nous avons demandé à des juges francophones de juger leur degré d'intelligibilité. Nous avons choisi un texte court et neutre : un paragraphe sur le Panthéon à Paris écrit pour un guide touristique (Appendice B) (Gralon, 2012).

Après cette première étape de l'évaluation des sept étudiantes étrangères, nous avons dû recruter une huitième étudiante pour refaire le protocole (tableau 2) car les juges francophones ne parvenaient pas à dégager trois étudiantes, dont le degré d'intelligibilité était jugé similaire, parmi les sept rassemblées pour la première étape. Après l'ajoute de cette huitième étudiante, nous avons pu dégager trois étudiantes d'origines ethniques différentes dont le degré d'intelligibilité était jugé semblable.

Tableau 2 : Présélection des étudiantes étrangères.

Identité	**Lieu de naissance**	**Langue maternelle**	**Date de naissance**	**Genre**	**Niveau de français**
Voix 1	Suisse	Allemand	1980	F	DALF C1
Voix 2	Taïwan	Mandarin et Taïwanais	1986	F	DELF B2
Voix 3	Syrie	Arabe	1986	F	DELF B2
Voix 4	Chine	Mandarin	1983	F	DELF B2-C1
Voix 5	Allemagne	Allemand	1988	F	DELF B2
Voix 6	Arabie Saoudite	Arabe	1989	F	Débutante
Voix 7	France	Arabe	1990	F	Maîtrise
Voix 8	Brésil	Portugais	1985	F	DELF C2

La tranche d'âge de ces huit étudiantes est de 22 à 32 ans. Toutes sont des étudiantes étrangères qui font leur Master à Grenoble, soit à l'Université Stendhal, soit à l'École de Commerce. Cinq d'entre elles ont un niveau de français comparable (DELF B2-DALF C1) tandis que deux sont des "outliers" (une plus forte et une plus faible que la majorité des autres). Compte tenu de l'origine de la rédactrice de ce mémoire, Canada, nous avons évité les accents anglophones. Comme nous l'avons évoqué en section 2.3.2, l'accent de l'expérimentateur peut influencer la perception des juges.

Le protocole s'est déroulé de la manière suivante. D'abord nous avons crée trois clips audio contenant les sept premiers extraits dans trois ordres aléatoires (1234567, 3571642, et 7462531). Ensuite, nous avons envoyé l'un ou l'autre de ces clips avec un court questionnaire (Appendice A) à vingt-quatre locuteur français natifs. Les locuteurs devaient écouter deux fois le clip et attribuer un des quatre niveaux d'intelligibilité pour chacune des sept étudiantes :

1 : Intelligibilité parfaite, compréhension aussi facile que si j'écoutais un français

2 : Très bonne intelligibilité, facile à comprendre

3 : Moyenne, je peux comprendre tout mais avec un peu d'effort

4 : Peu intelligible, assez difficile à comprendre

Nous avons saisi leurs réponses et nous avons fait des comparaisons des moyennes opposant deux à deux les jugements de chaque étudiante en utilisant des tests T appariées (tableau 3). Le but était de trouver trois étudiantes entre lesquelles les jugements des francophones ne faisaient apparaître aucune différence d'intelligibilité.

Après avoir réalisé ce premier travail sur les sept étudiantes impliquées dans la première étape, nous avons pu identifier deux paires de sujets pour lesquelles il n'y avait pas de différence statistiquement significative : l'étudiante taïwanaise et l'étudiante syrienne, moyenne de la première 2.46, moyenne de la deuxième 2.30; l'étudiante suisse et l'étudiante allemande, moyenne de la première 3.25, moyenne de la deuxième 3.17 (tableau 3).

Nous avons décidé de se concentrer sur la paire d'étudiantes syrienne ("arabe") et taïwanaise ("asiatique"), cherchant une étudiante "blanche" pour se joindre à elles. Nous avons recruté une étudiante brésilienne pour refaire le protocole. Dans le but de nous concentrer sur la paire d'étudiantes syrienne et taïwanaise, nous avons réduit le nombre total de possibilités à cinq, gardant l'étudiante syrienne et l'étudiante taïwanaise, ajoutant l'étudiante brésilienne, et choisissant deux autres étudiantes de manière aléatoire.

Après avoir fait passer de nouveaux les clips-audio auprès de dix-sept locuteurs français natifs, nous avons encore fait une comparaison des moyennes (tableau 4). Cette fois-ci, nous avons pu regrouper trois étudiantes ensemble entre lesquelles les moyennes des jugements sur leur intelligibilité n'étaient pas statistiquement différentes : l'étudiante syrienne, l'étudiante taïwanaise, et l'étudiante brésilienne.

Tableau 3 : comparaison de l'intelligibilité : présélection des étudiantes étrangères (étape 1)

Étudiante	1	2	3	4	5	6	7
Nationalité	Suisse	Taïwan	Syrie	Chine	Allemagne	Arabie Saoudite	France
Moyenne	3.25	2.46	2.30	2.75	3.17	1.04	3.96
Valeur t (valeur p)							
1 Suisse		5.60 (.0001)	10.46 (.0001)	4.30 (.0002)	0.97 (.3432)	19.15 (.0001)	-5.82 (.0001)
2 Taïwan			1.46 (.1581)	-2.74 (.0115)	-5.84 (.0001)	14.35 (.0001)	-15.02 (.0001)
3 Syrie				-4.56 (.0001)	-9.97 (.0001)	14.46 (.0001)	-17.64 (.0001)
4 Chine					-4.25 (.0003)	18.81 (.0001)	-12.12 (.0001)
5 Allemagne						24.22 (.0001)	-9.71 (.0001)
6 Arabie-Saoudite							-52.77 (.0001)

Tableau 4 : comparaison de l'intelligibilité : présélection des étudiantes étrangères (étape 2)

Étudiante	1	2	3	7	8
Nationalité	Suisse	Taïwan	Syrie	France	Brésil
Moyenne	3.29	2.18	2.18	4.00	2.24
Valeur t (valeur p)					
1 Suisse		9.500 (.0001)	9.500 (.0001)	-6.197 (.0001)	10.182 (.0001)
2 Taïwan			*moyennes identiques	-19.134 (.0001)	1.000 (.3322)
3 Syrie				19.134 (.0001)	.566 (0.579)
7 France					-16.641 (.0001)
8 Brésil					

Nous voulions trouver une manière efficace et claire de nommer les trois étudiantes enregistrés dans ce travail. Nous avons décidé de designer les étudiantes par le stéréotype qu'elles sont censés représenter. Comme nous l'avons évoqué en 1.2, nous ciblons les quatre stéréotypes identifiés par Simon et Clément (2006) : "arabe", "blanche", "asiatique" et "noir". Ainsi, nous avons choisi une lettre correspondant à un de ces stéréotypes pour chaque étudiante.

- l'étudiante syrienne est "l'étudiante A", désignant le stéréotype "arabe".
- l'étudiante brésilienne est "l'étudiante B", désignant le stéréotype "blanc".
- l'étudiante taïwanaise est "l'étudiante C", désignant le stéréotype "chinois". Puisque nous avons déjà utilisé la lettre "A" pour "arabe", nous avons décidé d'utiliser le terme "chinoise" pour remplacer le terme "asiatique".

Origine	**Code**	**Stéréotype**
Syrie	Étudiante A	A = "arabe"
Brésil	Étudiante B	B = "blanche"
Taïwan	Étudiante C	C = "chinoise"

2.2 Modification du texte à juger

Après avoir sélectionné les trois étudiantes, nous avons procédé à la création d'un texte contenant des erreurs qui devait être lues par ces dernières. Nous sommes partie du texte utilisé dans l'étape de présélection et nous avons créé artificiellement des erreurs susceptible d'être produites par tous les apprenants, quelque soit leur origine. Bien que la structure de la langue maternelle des apprenants ait un impact sur les erreurs grammaticales qu'ils commettent, "l'observation du discours en français des locuteurs d'origine étrangère montre que – quelle que soit leur langue d'origine - ils rencontrent aussi certaines difficultés récurrentes" (Desvaux, 2005 : 223). Desvaux remarque que les erreurs lexicales, qui peuvent dénoter de la provenance des locuteurs, sont statistiquement rares par rapport aux autre

erreurs.

Pour créer des erreurs vraisemblables, nous avons consulté trois bases de données qui présentent des corpus oraux d'apprenants de français des origines linguistiques diverses :

1. FLLOC (French Learner Language Oral Corpora) de l'Université d'Essex et de l'Université de Southampton en Angleterre. Ce corpus réunit des données produites par des apprenants de langues maternelles différentes et est accessibles gratuitement en ligne. On peut écouter les apprenants et ensuite lire des transcriptions pour identifier et classer les erreurs grammaticales.

2. Double Je. Ce corpus a été crée pour une étude de Desvaux (2005) qui s'appelle "L'asymptote du français avancé : les difficultés résistantes". Elle a transcrit les paroles des invités de l'émission française *Double je* entre 2004 et 2005. Les invités de cette émission, tous les étrangers, parlent un français avancé avec des "noyaux résistants".

3. Des étudiantes étrangères à l'Université Stendhal. Ce corpus a été crée par la rédactrice de ce mémoire. Les trois étudiantes étrangères choisies pour ce travail ont été enregistrées en train de discuter entre elles pendant quinze minutes. Nous avons écouté ces enregistrements, noté les erreurs grammaticales, et les avons classées sur des critères morphosyntaxiques.

Après avoir consulté ces trois corpus, nous avons identifié et classé certains types d'erreurs plus communes pour l'ensemble des apprenants :

1. les déterminants : "un vieux épave" (Double Je)

2. l'accord des adjectifs : "la société normatif" (Double Je)

3. les prépositions : "si on a choisi à vivre" (Étudiante à Stendhal)

4. les formes verbales :

   a) le subjonctif : "Bien qu'il est difficile" (FLLOC)

   b) les verbes pronominaux : "je ne peux pas arrêter de penser à ma famille" (Étudiante à Stendhal)

Nous voulions créer un texte court avec un densité d'erreurs suffisamment importante pour qu'elles ne soient pas trop faciles à trouver. En effet, il était nécessaire que les juges francophones auxquels nous demandions de chercher les erreurs puissent se différencier quant au nombre d'erreurs trouvées. En outre, nous avions besoin d'erreurs qui puissent être remarquées à l'oral. Dans une première étape, nous avons inséré cinq erreurs au texte sur le Panthéon utilisé dans l'étape de présélection des étudiantes étrangères (Appendice B). Toutefois, après avoir fait quelques essais auprès des Français de notre réseau de connaissance, nous avons perçu qu'ils trouvaient généralement toutes les erreurs. Ainsi, nous avons décidé d'augmenter le nombre à neuf. Le texte, avant et après sa version finale, se trouve ci-dessous avec les erreurs notées en caractères gras:

Texte sans erreurs :

Le Panthéon est l'un des monuments les plus célèbres de Paris. Il est situé sur la montagne Sainte-Geneviève dans le 5e arrondissement.

Ce monument est très grand et imposant. Il fait en tout 83 mètres de haut. Il a un dôme majestueux et trois coupoles. À l'intérieur, il y a une crypte et aussi des tableaux et des fresques illustrant la vie de sainte Geneviève. Bien qu'il soit exposé au soleil, il n'y a pas beaucoup de fenêtres et il fait donc froid et sombre à l'intérieur.

En 1885, le gouvernement a décidé de rénover le Panthéon. A partir du moment où les travaux ont été terminés, il est devenu un temple laïc destiné à honorer les français célèbres et à se souvenir des événements marquants de l'histoire de France.

Texte avec erreurs :

Le Panthéon est l'un des monuments les plus célèbres de Paris. Il est situé sur la montagne Sainte-Geneviève ***à le*** *5e arrondissement.*

Cette *monument est très grand et imposant. Il* ***est*** *en tout 83 mètres de haut. Il a un dôme* ***majestueuse*** *et trois coupoles. À l'intérieur, il y a une crypte et aussi des tableaux et des fresques illustrant la vie de sainte Geneviève. Bien qu'il* ***est*** *exposé au soleil, il n'y a pas beaucoup de fenêtres et il* ***est*** *donc froid et sombre à l'intérieur.*

En 1885, le gouvernement a décidé ***à*** *rénover le Panthéon. A partir du moment où les travaux* ***a*** *été terminés, il est devenu un temple laïc destiné à honorer les français célèbres et à* ***souvenir*** *des événements marquants de l'histoire de France*

Code	Erreur	Corrigée	Type d'erreur (selon le classement ci-dessus)
E1	À le 5e arrondissement	Au 5e arrondissement	1) déterminant 3)préposition
E2	Cette monument	Ce monument	1) déterminant
E3	Il est en tout 83 mètres de haut	Il fait en tout 83 mètres de haut	4) formes verbales
E4	Un dôme majestueuse	Un dôme majestueux	2) accord des adjectifs
E5	Bien qu'il est exposé au soleil	Bien qu'il soit exposé au soleil	4a) la subjonctif
E6	Il est donc froid et sombre	Il fait donc froid et sombre	4) formes verbales
E7	Décidé à rénover	Décider de rénover	3) prépositions
E8	Les travaux a été terminés	Les travaux ont été terminés	4) formes verbales
E9	Et à souvenir des événements	Et à se souvenir des événements	4b) verbes pronominaux

2.3 Enregistrement des versions finales du texte à juger

Nous avons effectué l'enregistrement du texte final, avec des erreurs, par les trois étudiantes, dans chambre sourde, au Gipsa Lab à l'Université Stendhal de Grenoble. Nous avons utilisé un enregistreur Zoom. La longueur des trois clips enregistrés était très semblables : étudiante A, 52 secondes ; étudiante B, 55 secondes ; étudiante C 53 secondes.

La dernière étape a consisté à noter des irrégularités apparues pendant les enregistrements qui ne figuraient pas parmi les erreurs insérées. Nous voulions prendre en compte ces erreurs ajoutées lors du traitement des données. Trois étudiantes françaises en Master 2 FLE (Pro) ont écouté les clips plusieurs fois et ont noté les aspects qui pourraient être vus comme une erreur (tableau 5).

Tableau 5 : Irrégularités des enregistrements finaux des trois étudiantes étrangères

Étudiante	Phonétique
A	- prononciation du mot "sainte" - prononciation du mot "situer"
B	- 'le' et 'de' prononcés comme 'les' et 'des' - le '8' en 1885 est mal prononcé - le mot 'moment' – pas de son nasal - Un petit bégaiement sur le 'bien'
C	- confusion entre le 'p' et le 'b' - "À la" 5ième au lieu de "à le" 5ième

3. Déroulement des expériences

3.1 : Recrutement des juges francophones

Une fois les trois enregistrements réalisés, nous avons préparé le déroulement des passations pour cueillir des jugements après des locuteurs francophone. Deux variables organisent la passation du protocole : trois *groupes de traitement* et deux *conditions* pour chacun de ces groupes (tableau 6). Chaque groupe de traitement écoutait l'enregistrement d'une des étudiantes – étudiante A, B ou C. Par ailleurs, chaque groupe de traitement se divisait encore en deux conditions. Dans la première condition, *photo*, nous fournissions une photo de l'étudiante étrangère aux juges et nous leur demandions après la passation d'estimer l'origine géographique de cette étudiante. Nous voulions savoir quelle était l'origine géographique attribuée par les juges à chaque étudiante et quel pouvait être l'effet de cette attribution sur les évaluations en langue étrangère. Dans la deuxième condition, *origine*, nous avons mentionné explicitement la langue maternelle de l'étudiante : "arabe", "portugaise" ou "chinoise". Nous voulions activer les trois stéréotypes en parallèle (arabe/blanc/asiatique).

Tableau 6 : les six conditions de passation différentes du protocole

Groupe de traitement	**Condition**
Bande sonore de l'étudiante A (syrienne)	1. Photo
	2. Langue "arabe"
Bande sonore de l'étudiante B (brésilienne)	3. Photo
	4. Langue "portugaise"
Bande sonore de l'étudiante C (taïwanaise)	5. Photo
	6. Langue "chinoise"

Le fait d'avoir ces six variantes différentes du protocole est motivé par des considérations issus des études sur la perception sociolinguistique, mentionné en 2.3. En effet, nous avons estimé très important de considérer l'aspect imaginaire de la perception. Nous voulions non seulement comparer les évaluations portant sur les trois étudiantes, mais nous voulions également examiner dans quelle mesure les évaluations d'une étudiante particulière dépendaient du fait que son origine était connu ou seulement "imaginé". Allions-nous constater des différences de jugement dans les cas où les juges écoutaient les mêmes énoncées et les mêmes erreurs, mais imaginaient les origines différents pour l'étudiante enregistrée? Notre étude étant exploratoire, nous pensions que ces six variantes permettraient d'aborder plusieurs aspects influençant les évaluations des juges.

3.2 Développement d'un questionnaire

À chaque juge francophone, nous avons proposé un questionnaire de deux pages, à remplir pendant et après l'audition des enregistrements des trois étudiantes (Appendice C). Au recto, la question consistait à demander aux juges de noter les fautes grammaticales qu'ils entendaient. On précisait que seules les fautes grammaticales étaient ciblées et qu'il fallait ignorer les erreurs de prononciation.

Consignes :
Vous allez écouter un clip audio d'une apprenante de français. Nous faisons une expérience pour savoir si l'évaluation de son français est la même quand elle est faite par des profs de français ou par d'autres personnes.
Vous n'allez entendre le clip audio qu'une seule fois. Pendant que vous écoutez, notez les erreurs grammaticales que vous repérez dans le tableau ci-dessous.
Attention:

1. *On cherche des erreurs grammaticales (conjugaison de verbes, genre, temps, etc.). À ne pas confondre avec des erreurs de prononciation.*

- *Il n'y a pas forcement 10 erreurs*
- *Nous vous remercions d'écrire de manière lisible*

Au verso, on demandait aux juges de faire trois évaluations globales en choisissant un des quatre points sur une échelle de Likert *(tout à fait d'accord/D'accord/Pas d'accord/Pas du tout d'accord).* Les trois questions correspondaient à trois dimensions : une dimension de maîtrise de la langue, une dimension d'intelligibilité de français, et une dimension de potentiel de performance scolaire dans le cadre de l'université. Ces trois évaluations globales sont reprises des critères pour l'examen oral du DELF niveau B2 (Appendice D).

Question	Tout à fait d'accord	D'accord	Pas d'accord	Pas du tout d'accord
1. Cette étudiante étrangère démontre une maîtrise de la langue qui lui permet de bien s'exprimer.	1	2	3	4
2. Malgré son accent, cette étudiante étrangère possède une prononciation intelligible qui n'entraîne pas de problèmes de compréhension.	1	2	3	4
3. Selon sa performance dans le clip audio, cette étudiante étrangère serait capable de suivre un cursus universitaire en France.	1	2	3	4

La dernière section demandait aux juges leur langue(s) maternelle(s), leur pays et date de naissance, les langues parlées par leurs parents, et les autres langues qu'ils connaissaient. Nous avons également demandé s'ils ont vécu plus de trois mois dans un autre pays que la France. Ces informations biographique nous aidaient à éliminer certains juges lors de la saisie des données.

Les juges auxquels l'origine ethnique de l'étudiante n'étaient pas mentionné ont reçu des questionnaires identiques aux autres. Ces questionnaires comportait une petite photo (5cm x 7cm) du visage de l'étudiante[10]. Ces photos étaient prises par le rédactrice de ce mémoire. Chaque étudiante adopte une pose similaire, porte des vêtements quelconques et a une expression neutre. La toile de fond est la même pour toutes les trois. En outre, les juges qui ne disposaient pas de la nationalité des

10 Les photos ne sont pas mises en appendice pour préserver l'anonymat des étudiantes

étudiantes mais seulement de leur photo ont rempli une troisième section du questionnaire, ou figurait une carte du monde. Nous demandions alors aux juges d'indiquer le pays ou la zone géographique dont était susceptible de venir l'étudiante. Ces réponses nous ont permis de savoir si le juge pouvait identifier correctement l'origine de l'étudiante et quelle représentation il ou elle avait en tête au sujet de sa nationalité.

3.3 Déroulement

Au cours de deux semaines (décembre, 2012), nous avons recueilli 343 questionnaires auprès d'étudiants francophones rencontrés dans douze cours différents à l'Université Pierre Mendès France et à l'Université Stendhal (Tableau 7).

Nous nous sommes assurée de suivre la mêmes procédure avec chaque groupe. Après avoir donné les questionnaires, nous présentions une courte explicitation du projet aux juges (Appendice E). Afin d'éviter les biais, cette introduction attribuait à l'étude un objectif différent de son but réel. Nous expliquions que le but de notre recherche concernant des apprenants du français à Grenoble, était de savoir si l'évaluation du français des apprenants était la même quand elle était faite par des professeurs de français ou par des francophones non-professionnel de l'enseignement.

Dans la condition ou la nationalité de l'étudiante était donnée aux juges, nous remplacions "une étudiante étrangère" avec "une étudiante étrangère arabe/chinoise/brésilienne".

Tableau 7 : Liste des passations des protocoles

#	Groupe de Traitement	Condition	École	Cours	Niveau	# question-naires total	Éliminé : pas de locuteur natif	Éliminé : pas complété	Éliminé : Parle la langue ciblé
1	B	photo	UPMF	Statistique	L2	18	3	0	0
2	C	photo	UPMF	Statistique	L2	24	2	0	0
3	A	photo	UPMF	Psychologie	L1	75	3	1	3
4	A	photo	Stendhal	Statistique	L2	12	1	0	1
5	A	origine	UMPF	Psychologie	L1	103	4	1	5
6	A	photo	Stendhal	Statistique	L2	13	0	0	0
7	B	photo	UPMF	Histoire	L2	21	0	0	0
8	B	origine	UPMF	Histoire	L2	11	1	0	0
9	C	origine	UPMF	Histoire	L2	16	0	0	0
10	B	photo	UPMF	Histoire	L2	14	0	0	1
11	B	origine	UPMF	Histoire	L2	18	0	0	0
12	B	origine	UPMF	Histoire	M1	18	3	0	1
						343	17	2	11

	A photo	A origine	B photo	B origine	C photo	C origine
Total # de juges	91	93	50	42	22	16

III. PRÉSENTATION ET ANALYSE DES RÉSULTATS

1.1 La saisie des données

Après avoir fini le recueil de données, nous avons éliminé les juges qui ne s'étaient pas désignés comme des locuteurs natifs de français, ainsi que les questionnaires incomplets. En outre, nous avons éliminé les juges qui parlaient la langue maternelle de l'étudiante qu'ils écoutaient : groupe A, arabe, y compris ses dialectes; groupe B, portugais; groupe C, mandarin, cantonais ou un dialecte de chinois. En effet, comme nous l'avons évoqué en 2.5, les gens sont parfois plus indulgents envers des locuteurs dont ils parlent la langue (Winke, Gass, & Myford, 2011). Après cette sélection, il nous restaient 313 questionnaires (tableau 7).

Nous avons catégorisé chaque sujet dans un fichier en utilisant les variables suivants:

1. Condition : A Photo, A Origine, B Photo, B Origine, C Photo, et C Origine
2. Erreurs notées : 0 ou 1 pour chacune des neuf erreurs insérées au texte
3. Erreurs totales : valeur entre 0-9
3. 'Erreurs supplémentaires" : erreurs qui ne figuraient pas sur le clip : notées exactement comme les ont écrites les juges
4. Évaluations globales : valeur entre 1-4 selon la réponse sur l'échelle de Likert pour les trois questions holistiques (maîtrise de la langue, intelligibilité, potentiel académique)
5. Origine estimé (dans le cas de la carte) : zone géographique codée selon les quatre stéréotypes de Simon et Clément (2006) : 'arabe', 'noire', 'asiatique', et 'blanc'. Asiatique (1) noir (2) arabe (3) ou blanche (4).[11] (Appendice F)
6. Date de naissance
7. Pays de naissance
8. Langue maternelle
9. Langue(s) parlé(es) par les parents
10. S'ils ont vécu plus de trois mois dans un autre pays que la France : valeur de 0 ou 1

Après la saisie, nous avons importé ces données dans le logiciel SPSS pour faire des analyses statistiques. Une description détaillée de ces analyses se trouvent dans les lignes suivantes.

11 Certains réponses étaient plus faciles à classer que d'autres. Quand une réponse était ambiguë, nous avons demandé auprès des Français de notre réseau de connaissance pour savoir le stéréotype dominant pour cette zone.

2. Le traitement des données

L'analyse statistique des données recueillies des 313 questionnaires se réalisera suivant trois étapes. Les deux premières étapes aborderont les comparaisons des trois *groupes de traitement* (juges écoutant les clips audio de l'étudiante A, l'étudiante B et l'étudiante C). Dans une première étape, il s'agira d'une comparaison des erreurs grammaticales notées entre les trois groupes de traitement. Dans une seconde étape, nous mettrons les résultats des erreurs en relation avec trois jugements globaux (*maîtrise de la langue, intelligibilité, potentiel académique).* Nous chercherons un rapport entre le nombre d'erreurs trouvées et les scores globales dans les trois groupes de traitement.

Dans la dernière étape, nous nous interrogerons sure l'influence du stéréotype "arabe" sur les jugements d'un même enregistrement. Dans ce but, nous opposerons les évaluations des juges qui ont été informés explicitement de l'origine "arabe" de l'étudiante syrienne (tous les juges de la condition *origine*) à ceux qui ont attribué une origine non-arabe sur le questionnaire (certains juges de la condition *photo*). Ainsi, nous arriverons à comparer les évaluations d'un même enregistrement par les juges ayant des croyances différentes vis à vis de l'origine du locuteur.

Tableau 8 : Résumé des trois analyses

<table>
<tr><th>groupe de traitement</th><th>Condition (# de juges)</th><th>Analyse #1 : erreurs par groupe de traitement</th><th>Analyse #2 : évaluations holistique par groupe de traitement</th><th>Analyse #3 : par condition</th></tr>
<tr><td rowspan="2">Étudiante A (arabe)</td><td>Photo (91)</td><td rowspan="2">A
vs</td><td rowspan="2">A
vs</td><td>origine "arabe" non-connue (réponse incorrecte sur la partie zone géographique)
vs</td></tr>
<tr><td>Origine (93)</td><td>origine connue (explicitée par l'expérimentateur)</td></tr>
<tr><td rowspan="2">Étudiante B (brésilienne)</td><td>Photo (49)</td><td rowspan="2">B
vs</td><td rowspan="2">B
vs</td><td rowspan="4">N/a</td></tr>
<tr><td>Origine (42)</td></tr>
<tr><td rowspan="2">Étudiante C (chinoise)</td><td>Photo (22)</td><td rowspan="2">C</td><td rowspan="2">C</td></tr>
<tr><td>Origine (16)</td></tr>
</table>

Ces trois analyses statistiques nous permettront d'identifier et de quantifier les différences qui existent entre les évaluations des étudiantes différentes ainsi qu'entre les évaluations d'un même étudiante par les juges ayant les croyances différentes sur son identité. À travers des tests de significativité, nos analyses nous permettront de savoir si les différences éventuelles entre les moyennes sont dues au hasard ou si elles sont susceptibles à correspondre à des tendances systématiques dans la population. À l'issue de ces analyses, nous ferons une synthèse des résultats, en nous interrogeant sur la validité écologique des différences statistiquement significatives constatées dans l'expérience. Plus précisément, nous demanderons si ces différences ont une signification en dehors du contexte expérimental.

3. Analyse #1 : Erreurs grammaticales par groupe de traitement

3.1 Tendances générales des erreurs

Pendant l'écoute du clip audio, les juges ont noté les erreurs grammaticales qu'ils ont entendues. Dans leur ensemble, le nombre d'erreurs noté par chaque juge suit une distribution normale (figure 1), avec une moyenne de 3.18, une médiane de 3, et un mode de 3. Le maximum est 7 et le minimum est 0. Personne n'a trouvé les neufs erreurs possibles, ce qui s'explique probablement par le fait que les juges n'ont entendu le clip audio qu'une seule fois. Certaines erreurs sont très rapprochées dans le temps; il est probable que le fait de noter une erreur leur en a fait manquer d'autres.

Afin d'analyser les moyennes des totales des erreurs par groupe de traitement, nous avons fait un ANOVA suivi par des tests T (tableau 9). Il existe entre les trois groupes une différence statistiquement significative ($p = 0.003$); cette différence se trouve entre le groupe "chinois" et "arabe" ($p=0.010$) et entre le groupe "chinois" et le groupe "blanc" ($p=0.002$) (tableau 10).

Ainsi, l'étudiante taïwanaise se distingue des deux autres du fait que ses fautes sont en général plus remarquées par les juges. Par ailleurs, la médiane de ce groupe est élevée par rapport aux autres : elle

est à 4 au lieu de 3. C'est ensemble de différences suggèrent que les juges ont trouvé plus facilement les erreurs dans l'enregistrement de l'étudiante présentant le stéréotype "chinois".

Figure 1 : La distribution du total des erreurs notées

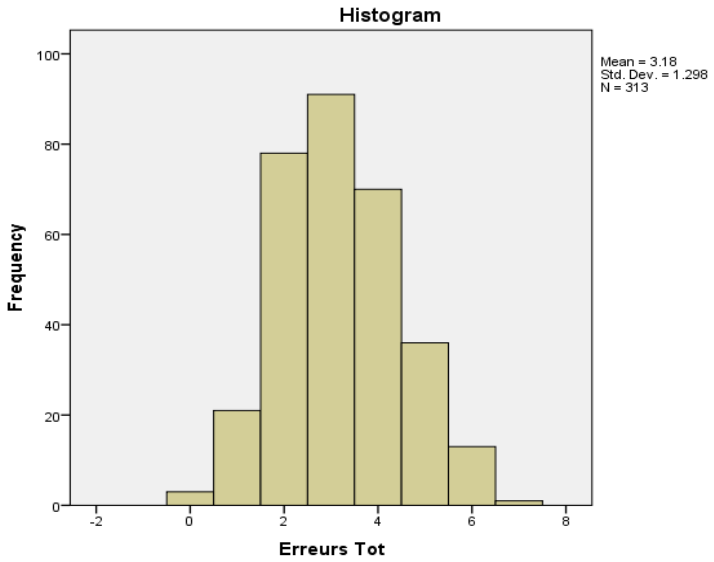

Tableau 9 : totales des erreurs par groupe de traitement

	moyenne	médiane	variance	Écart type	Min-Max
A	3.15	3.00	1.711	1.308	0-7
B	2.98	3.00	1.622	1.273	1-6
C	3.82	4.00	1.289	1.136	1-6

Tableau 10 : Comparaisons des paires des groupes de traitement : totales des erreurs

(I) Treatment group	(J) Treatment group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval
					Lower Bound
A	C	-.669*	.228	.010	-1.21
	B	.169	.164	.559	-.22
C	A	.669*	.228	.010	.13
	B	.838*	.247	.002	.26
B	A	-.169	.164	.559	-.55
	C	-.838*	.247	.002	-1.42

3.2 Analyses des Erreurs Une par Une

Dans l'ensemble, certains erreurs étaient plus remarquées que d'autres, notamment E1 et E7 (Tableau 11). Ce phénomène peut s'expliquer par le fait que E1 est la première erreur et E7 se trouve après une pose naturelle. E2 et E9, aussi fréquemment notées, sont près du début et de la fin du discours. Les erreurs les moins notées étaient E4, E5 et E6. Elles se trouvent au milieu d'une période prolongée de parole. Il est possible que ces tendances générales soient dû au fait que l'on a tendance à mieux remarquer et rappeler les éléments d'une liste qui se trouvent au début et à la fin (Murdock, 1962).

Bien qu'il y ait des tendances générales, il existe des différences marquées entre les trois groupes de traitement. Afin de savoir si ces différences entre groupes sont statistiquement significatives, nous avons fait des tests de chi-deux afin de comparer dans chacune des trois conditions la proportion de juges qui ont détecté l'erreur. Nous avons relevé une différence statistiquement significative entre les groupes de traitement pour six des neufs erreurs (lignes grises du tableau 11). E6/E8 sont plus remarquées chez l'étudiante C, E4 est plus remarquée chez l'étudiante B et E2/E9 sont plus remarquées chez l'étudiante A. Dans la partie discussion, nous mettrons en relation ces résultats avec ceux de la comparaison des moyennes.

Tableau 11 : % des juges qui ont trouvé les erreurs individuelles. Total et par groupe de traitement

Code	Erreur	Total (A+B+C)	Groupe A	Groupe B	Groupe C	Valeur p
E1	À le 5e arrondissement	76.6	72.8	78.9	89.5	.073
E2	Cette monument	34.5	**43.5**	24.2	15.8	.000
E3	Il est en tout 83 mètres de haut	27.8	14.1	8.3	5.4	.033
E4	Un dôme majestueuse	20.8	15.2	**34.1**	15.8	.001
E5	Bien qu'il est exposé au soleil	15.3	12.0	18.7	23.7	.108
E6	Il est donc froid et sombre	13.1	5.4	8.8	**60.5**	.000
E7	Décidé à rénover	63.9	63.0	70.3	52.6	.151
E8	Les travaux a été terminés	32.3	38.0	12.1	**52.6**	.000
E9	Et à souvenir des événements	33.9	**40.8**	23.1	26.3	.008

Il est intéressant de noter que certaines fautes remarquées par les juges sont des "erreurs supplémentaires" - elles ne figurent pas sur la liste des neufs erreurs intentionnellement insérées au texte. Nous avons noté trente-et-une erreurs supplémentaires à travers tous les questionnaires, ce qui représente de 5.8 à 10.8% de toutes les erreurs marquées selon le groupe de traitement. Nous avons noté ces "fautes" exactement comme les juges les ont écrites (tableau 12), mais nous n'avons pas classé ou analyser ces réponses. La plupart d'entre elles sont les mots ou les courtes expressions, comme par exemple "crypte" et "elle est". Nous supposons que les juges entendaient une prononciation non-standard de ces mots, interprétant cette prononciation comme une faute. Cela indique une confusion de la part des juges sur la distinction entre une erreur grammaticale et une prononciation non-standard ou aberrante. Nous avons eu de la difficulté à placer la frontière entre ces deux concepts afin de les analyser. De plus, nous ne savons pas avec certitude si ce phénomène s'explique du fait que

certains juges n'ont pas entendu ou compris les instruction ou du fait qu'ils n'ont simplement pas su faire la différence entre une faute de grammaire et une prononciation non-standard. Il serait intéressant de poursuivre cette questions dans de futures recherches.

Tableau 12 : Des erreurs supplémentaires (ES) par groupe de traitement

Groupe de traitement	# ES	ES : % du total d'erreurs
A	73	10.8%
B	21	7.0%
C	9	5.8%

Liste des ES

1. renouvelé 2. fresque 3. crypte 4. situé 5. les plus marquants 6. les Panthéon 7. illustrant 8. Paris 9. dôme 10. pas de beaucoup 11. banque	12. gouvernement 13. à partir du moment 14. il est (au lieu de) il a 15. n'est pas (au lieu de) n'a pas 16. grand 17. l'un des monuments 18. elle est 19. il est devenu 20. "i" "a" "de"	21. temple 22. de France 23. intérieur 24. ville de St Janvier 25. elle a oublié "la montagne" 26. en temps 27. destiné à honorer 28. la vie de St. Geneviève 29. au 5e siècle 30. toit coupole 31. et aussi	

3.3 Discussion sur la détection des erreurs

Rappel de question de recherche #1 :
Comparaison 1 : entre trois étudiantes étrangères lisant un même texte comportant les mêmes erreurs grammaticales :
1) est-ce que les juges trouveront le même nombre total d'erreurs et les mêmes erreurs chez chaque étudiante?

Quand trois étudiantes étrangères lisent un même texte comportant des erreurs grammaticales, les étudiants français ne repèrent ni le même nombre, ni le même type d'erreur à travers les trois. Davantage d'erreurs sont notées pour le stéréotype "chinois" et moins avec le stéréotype "arabe", le stéréotype "blanc" se trouvant au milieu. Les différences entre "chinois"/"arabe" et "chinois"/"blanc" sont statistiquement significatives, ce qui nous permet de conclure que cette différence de moyenne ne

se produit pas par hasard. Pourtant, bien qu'en moyenne, les juges trouvent plus de fautes chez l'étudiante taïwanaise, ils ont plus tendance à trouver certaines erreurs (E2, E4, et E9) chez les deux autres étudiantes.

Nous avons sélectionné trois étudiantes étrangères ayant un niveau d'intelligibilité semblable, mais il s'avère que toutes les autres variables dans la parole (e.g. l'intonation, le débit, la prononciation) peuvent avoir un impact important sur la détection d'erreurs. Effectivement, ces résultats soulignent la difficulté à comparer les évaluations des enregistrements différents. Même ayant un niveau d'intelligibilité semblable et faisant les mêmes fautes, chaque locuteur a sa manière unique à prononcer le texte. Effectivement, ces petits différences entre les enregistrements peuvent avoir une grande influence sur quelles fautes sont réparées chez chaque locuteur.

Malgré la composition différente d'erreurs trouvées chez chaque étudiante, une différence significative existe entre les moyennes des totales. Enfin, les juges ont noté plus de fautes chez l'étudiante taïwanaise. Il est possible que l'étudiante taïwanaise ait prononcé les erreurs plus clairement que les deux autres mais nous n'avons aucun moyen pour examiner ce facteur supplémentaire.

Cette conclusion conduit logiquement à la question traitée dans la prochaine analyse : est-ce qu'il y a un rapport entre le nombre total d'erreurs notées et les évaluations globales de compétence en langue? Autrement dit, les juges, ayant trouvé plus de fautes chez l'étudiante taïwanaise, seront-ils plus sévères envers elle sur les questions globales qui porte sur la maîtrise de la langue, l'intelligibilité, et le potentiel académique?

4. Analyse #2 : Évaluations globales par groupe de traitement

Le recto du questionnaire demande aux juges de noter les erreurs concrètes de grammaire; le verso leur demande de faire trois évaluations globales portant sur les compétences globales de l'étudiante. Chaque réponse de 1 à 4 correspond à l'échelle Likert ci-dessous. Rappel :

Question	Tout à fait d'accord	D'accord	Pas d'accord	Pas du tout d'accord
1. Cette étudiante étrangère démontre une maîtrise de la langue qui lui permet de bien s'exprimer.	1	2	3	4
2. Malgré son accent, cette étudiante étrangère possède une prononciation intelligible qui n'entraîne pas de problèmes de compréhension.	1	2	3	4
3. Selon sa performance dans le clip audio, cette étudiante étrangère serait capable de suivre un cursus universitaire en France.	1	2	3	4

4.1 Tendances générales des évaluations globales

Dans l'ensemble, les juges se montraient réticents pour attribuer des notes de 3 et 4. Par conséquent, les moyennes se trouvent toutes entre 1 et 2 (tableau 13). On constate qu'à travers les trois questions, les juges donnent, en moyenne, de meilleures notes à l'étudiante brésilienne. Cette différence n'est statistiquement significative que pour la question 3 sur le potentiel académique (tableau 14). Dans ce cas, une différence existe entre A/B (p=0.007) et B/C (p= 0.046). On remarque très peu de différence entre les scores pour l'étudiante C (taïwanaise) et l'étudiante A (syrienne).

Tableau 13 : comparaison des moyennes des évaluations globales

groupes de traitement	Maîtrise	Intelligibilité	Université	Total
Étudiante A	1.72	1.80	1.76	5.28
Étudiante B	1.58	1.70	1.56	4.84
Étudiante C	1.74	1.87	1.76	5.37
Différence Significative?	non p=0.088	non p= 0.174	oui p = 0.015	

Tableau 14 : comparaison des moyennes, question 'université'

Groupes de traitement	Différence significative?	Valeur p
A/B	oui	0.007
A/C	non	0.990
B/C	oui	0.046

4.2 Discussion des évaluations globales

Rappel de question de recherche #2

2.) quel est le rapport entre les erreurs identifiées par les juges et les évaluations globales des compétences en langue étrangère attribuées par ces juges aux étudiantes?

Un problème auquel nous avons été confrontée pendant l'analyse précédente était l'étroite proximité des moyennes. En effet, toutes les moyennes sont très proches en terme de valeur mais aussi en terme de signification. Selon notre échelle de Likert, une note de '1' correspond à 'tout à fait d'accord' et une note de '2' correspond à 'd'accord'. Ainsi, il est difficile de prétendre que la différence entre 1.56 et 1.76 soit très importante quant à ce qu'elle représente dans le jugement des sujets. Les juges avaient des réponses, en général, favorables envers les trois étudiantes. Ce resserrement des notes vers le haut peut traduire l'influence du filtre social sur l'expression des préjugés. Par contre, il est possible qu'il n'existait pas suffisamment de points sur l'échelle de Likert pour étaler les réponses.

Nous pouvons néanmoins constater un rapport logique entre les deux éléments de jugement : quantitatif et qualitatif. Les juges ont trouvé plus d'erreurs chez l'étudiante C et lui ont accordé une note holistique plus basse que l'étudiante B. De même, les juges ont trouvé moins d'erreurs chez l'étudiante B et se sont montrés légèrement plus favorables lors des évaluations globales. Pourtant, les tendances ne sont pas très claires – par exemple, les juges ont noté moins d'erreurs chez l'étudiante A que chez l'étudiante C, mais ils l'ont accordée des scores globaux presque pareils. Soit il n'y a pas de rapport entre ces deux types d'évaluation soit les faibles différences entre les scores sur l'échelle de

Likert ne nous permettent pas de bien distinguer les tendances.

5. Analyse #3 : L'identité imaginée du locuteur

L'analyse par groupe de traitement étant terminée, nous allons considérer maintenant les différences en fonction de l'identité imaginée du locuteur. Plus précisément, nous allons comparer deux groupes de juges qui ont tous entendu la même bande sonore mais qui avaient des croyances différentes sur l'identité du locuteur. Le premier group se compose des juges qui, voyant une photo du visage de l'étudiante syrienne, lui ont attribué une zone géographique non-arabophone (voir 1.1). Le deuxième groupe se compose des juges qui ont été informés explicitement de l'identité "arabe" de l'étudiante (tableau 16). Ainsi, nous pouvons opposer les évaluations d'un même enregistrement entre ceux qui avaient un stéréotype "arabe" en tête et ceux qui n'en avaient pas. Cette analyse a été rendu possible parce que la plupart des juges (94%) qui écoutaient l'étudiante syrienne et voyaient sa photo ne lui ont pas attribué une identité "arabe".

Tableau 15 : les deux groupes de juges pour l'analyse 3

Groupe pour l'analyse 3	**Groupe de traitement**	**condition**	**# de juges dans le groupe**
Stéréotype "arabe"	Étudiante A	origine	93
Stéréotype "non-arabe"	Étudiante A	photo	85

5.1 Utilisation d'un "groupe témoin" pour savoir si un changement de protocole (origine explicitée vs. photo) peut avoir une influence sur les réponses des juges

Nous avons constaté dans les études (citées en 2.3.2 de la section "Approches théoriques") que les évaluations peuvent être influencées par de petits changements dans l'environnement. Dans l'analyse actuelle, nous voulions nous assurer que les différences éventuelles entre les deux groupes ("arabe" et "non-arabe") ne sont pas provoquées par le changement de protocole. En effet, les juges du groupe "non-arabe" disposaient d'une photo de l'étudiante et les juges du groupe "arabe" n'en avaient pas.

Nous avons donc décidé de faire une comparaison des juges des deux protocoles (photo et origine explicitée) pour l'étudiante brésilienne. Le cas de l'étudiante brésilienne se prête bien à ce rôle de "groupe témoin" parce que 99% des juges de cette étudiante dans la condition photo lui ont attribué une origine "blanche". Autrement dit, presque tous les juges dans les deux conditions pensaient que l'étudiante brésilienne venaient d'une zone géographique codée "blanche". Le fait de comparer les évaluations des deux conditions de l'étudiante brésilienne nous permettra de savoir si le changement de protocole (le fait de voir une photo ou non) aurait pu provoquer les différences de jugement.

Comparer les analyses des deux conditions de l'étudiante A (entre juges qui la perçoivent comme "arabe" et juges qui la perçoivent comme "non-arabe") *et* de l'étudiante B (où tous les juges la perçoivent comme "blanche") nous permettra de identifier ou éliminer *condition* comme variable qui peut avoir une influence significative sur les juges. Ainsi, nous pourrons dire avec davantage de certitude que les différences éventuelles entre les juges des deux groupes de l'étudiante "arabe" sont dues à la perception de son origine et non au changement de protocole.

5.2 Erreurs grammaticales

Nous avons fait un test de chi-deux pour savoir si les différences en moyenne du total d'erreurs étaient statistiquement significative. Dans le cas de l'étudiante A, les juges qui la percevaient comme "arabe" avaient tendance à trouver moins de fautes que les juges qui la percevaient comme "non-arabe" (tableau 17). Cette tendance est significative (p=0.000). Il n'y a pas de différence statistiquement significative entre ces deux conditions du groupe B. Cette tendance se répète dans les analyses des erreurs de cas par cas : les différences statistiquement significatives se trouvent entre les deux groupes de juges de l'étudiante A (E1 et E4) et non entre les deux groupes de juges de l'étudiante B.

Tableau 16 : Comparaison des erreurs grammaticales par condition : groupe A et groupe B

	Groupe A					**Groupe B**			
	"arabe"	**"non-arabe"**	**Sig?**	**Valeur p**		**Origine**	**Photo**	**Sig?**	**Valeur p**
Moyenne	2.80	3.51	Y	0.000		3.24	2.76	N	.071
E1	63.4	82.4	Y	.006		85.7	72.9	N	.220
E2	44.1	42.9	N	.985		26.2	22.4	N	.865
E3	21.5	26.4	N	.548		31.0	26.5	N	.816
E4	8.6	22.0	Y	.020		35.7	32.7	N	.932
E5	9.7	14.3	N	.462		21.4	16.3	N	.724
E6	4.3	6.6	N	.718		9.5	8.2	N	1.000
E7	62.4	63.7	N	.968		69.0	71.4	N	.986
E8	31.2	45.1	N	.074		16.7	8.2	N	.359
E9	34.4	47.3	N	.105		28.6	18.4	N	.367

5.3 Tendances générales des évaluations globales

Pour comparer les moyennes des évaluations globales, nous avons fait des tests T. Pour les deux groupes de juges de l'étudiante A, ceux qui connaissaient son origine "arabe" ont donné des scores moins favorables pour chacune des trois questions sur la compétence globale (tableau 18). Toutes ces trois différences sont statistiquement significatives. Par contre, on ne constate pas de telles différences entre les deux groupes de juges de l'étudiante B.

Les différences remarquées dans le groupe A semblent être en opposition avec le nombre d'erreurs notées. En effet, les juges qui connaissaient l'origine "arabe" de l'étudiante A, ont trouvé moins d'erreurs grammaticales mais l'ont jugé plus sévèrement lors des évaluations globales. Comme on ne remarque pas les différences entre les deux conditions du groupe B, nous pouvons dire que le fait de changer le protocole n'a pas provoqué ces différences particulières chez les juges dans le groupe A.

Tableau 17 : Évaluations globales par condition : groupe A et groupe B

groupe de traitement	Condition	Maîtrise	p	Intelligibilité	p	Université	p
Groupe A	"non-arabe"	1.59	0.001	1.65	0.000	1.67	0.032
	"arabe"	1.85		1.95		1.85	
Groupe B	Photo	1.60	0.711	1.68	0.700	1.56	0.99
	Origine	1.56		1.72		1.56	

5.4. Effet des conditions sur les évaluations globales : analyse des proportions

Comme nous l'avons évoqué dans l'analyse par groupe de traitement, la différence entre une moyenne de 1.59 et 1.85 n'est pas très importante en terme de signification – les deux chiffres représentent des évaluations favorables. La raison pour cette proximité des moyennes est le fait que peu de juges ont accordé des '3' (*pas d'accord*) et '4' (*pas du tout d'accord*) sur l'échelle de Likert. Pourtant, en faisant l'analyse sur l'étudiante A, nous avons avons remarqué une incidence importante de ces évaluations défavorables. Nous avons donc décidé de faire une analyse différente des évaluations globales afin de comparer la *proportion* des réponses favorables (*tout à fait d'accord/d'accord*) à celle des réponses défavorables (*pas d'accord/pas du tout d'accord*).

Nous avons d'abord calculé le pourcentage des réponses favorable et défavorable. Ensuite, nous avons appliqué le test exact de Fisher pour pouvoir comparer les proportions de réponses défavorables entre les deux groupes de juges (tableau 19). En effet, ce test nous permet de savoir si la distribution des réponses défavorables aurait pu se produire par hasard ou non.

Les seules comparaisons qui atteignent le seuil de signification statistique se trouvaient dans le groupe A. En effet, les juges qui connaissaient l'origine "arabe" de cette étudiante ont donné

significativement plus d'évaluations défavorables que les juges qui ne la connaissaient pas (tableau 19). Ce phénomène est présent à travers les trois questions (maîtrise, prononciation, potentiel académique). La différence n'est pas statistiquement significative pour la question sur le potentiel académique (p=0.09) alors qu'elle est pour les autres deux questions (p=0.035 et 0.001). Dans le groupe B, entre la condition 'photo' et 'origine', il existe aucune différence significative entre les réponses défavorables.

Tableau 18 : Des réponses défavorables sur les évaluations globales

groupe de traitement	**Question**	**"non-arabe" % (n)**	**"arabe" % (n)**	**Signifiant?**	**Valeur p**
A	Maîtrise	1.1% (1)	7.5% (7)	Y	.035
	Prononciation	1.1% (1)	12.9% (12)	Y	.001
	Université	2.2% (2)	7.5% (7)	N	.090
B	**Question**	**photo**	**origine**	**Signifiant?**	**Valeur p**
	Maîtrise	2.0% (1)	2.4% (1)	N	.713
	Prononciation	2.0% (1)	2.4% (1)	N	.713
	Université	4.1 (2)	7.1% (3)	N	.427

5.5 Discussion des résultats des analyses par origine "arabe" et "non-arabe"

Rappel de question de recherche #3 :
3.) est-ce que les évaluations d'une même étudiante étrangère seront différentes si son origine est perçue comme "arabe"?

À travers toutes les analyses, nous avons constaté des différences significatives entre les juges qui percevaient l'origine de l'étudiante A comme "arabe" et les juges qui la percevaient comme "non-arabe". Cette tendance ne peut pas être attribuée au changement de protocole car elle n'est pas présente entre les deux groupes de juges de l'étudiante B. Cela suggère que le fait de connaître l'origine "arabe" de l'étudiante A a eu un impact sur les réponses des juges. En effet, les juges, qui connaissaient l'origine

de cette étudiante, ont trouvé moins d'erreurs grammaticales, mais ils se sont montrés plus sévères lors des évaluations globales. Il est intéressant de noter que les seules évaluations du type '4' (*pas du tout d'accord)* de l'échelle de Likert, à travers tous les juges des trois étudiantes, se trouvent aussi chez ces juges du groupe "arabe" de l'étudiante A.

Les différences entre les conditions du groupe A ne sont pas très importantes mais sont néanmoins statistiquement significatives. Cela suggère que les différences ne se sont pas produites par hasard et que la variable d'origine "arabe" perçu a influencé les résultats. Nous pouvons confirmer que ce n'était pas le fait de changer le protocole qui a provoqué ces différences parce que l'on n'a pas remarqué de différences statistiquement significatives chez les juges des deux conditions du groupe B.

SYNTHÈSE DES RÉSULTATS

1. Résumé des résultats

Les analyses par groupe de traitement démontrent que les juges trouvent plus d'erreurs chez l'étudiante taïwanaise que chez les autres deux étudiantes et qu'ils accordent un meilleur score à l'étudiante brésilienne sur les évaluations globales. Ces résultats sont statistiquement significatives. On ne peut pas tirer de grandes conclusions de ces différences parce que chaque groupe de juges écoutait un enregistrement d'une étudiante différente, et dans chaque enregistrement l'étudiante étrangère avait sa manière particulière à prononcer le texte. Le grand nombre de variables que nous ne pouvions pas contrôler telles que l'intonation et le débit de chaque étudiante fait que les différences sur le nombre de fautes remarquées entre les groupes de traitement peuvent être attribuées à d'autres variables que l'origine de chaque étudiante. Nous avons que les juges repéraient plus de fautes chez l'étudiante taïwanaise, mais il est impossible de dire avec certitude si ce phénomène est causé par son identité "asiatique" ou si elle prononce les fautes de façon à les rendre plus facile à réparer. Enfin, nous devons reconnaître que ce protocole, comparant un même texte lu par trois étudiantes, rend très difficile le fait d'isoler une seule variable.

Toutefois, les conclusions tirées de notre troisième analyse sont plus nettes. Nous avons comparé les juges qui connaissaient l'origine "arabe" de l'étudiante A et les juges ne la connaissaient pas et nous avons révélé que cette connaissance a exercé une influence sur leurs jugements de la compétence langagière de cette étudiante. Comme tous ces juges écoutaient un même enregistrement, on évite le grand nombre de variables que l'on a vu dans les deux premières analyses. Afin d'éliminer le changement de protocole comme variable influente dans cette analyse de l'étudiante A, nous avons comparé des juges des deux conditions pour l'étudiante B, ce qui ne révélait pas de différences significatives. Nous pouvons donc conclure que l'origine "arabe" imaginée a joué un rôle dans les

évaluations des juges.

Dans l'ensemble, les juges qui connaissaient son origine "arabe" de l'étudiante A l'ont accordé un score global plus bas et plus d'évaluations défavorables. Ces résultats sont particulièrement intéressants étant donné que ces mêmes juges (groupe "arabe") ont trouvé *moins* de fautes dans son énoncée. Ce rapport, le contraire de celui qui était attendu, suggère que les évaluations globales de la compétence en langue ne sont pas liées directement aux mesures quantitatives, comme le nombre de fautes grammaticales, mais sont fondées sur en ensemble de critères, à la fois explicites et implicites.

Vu dans l'optique de l'automaticité, il semble fort possible que nous soyons susceptibles à fonder des jugements de compétence langagière sur des facteurs implicites ou des associations automatiques. Dans le cas de l'étudiante syrienne, les juges qui étaient informés explicitement de son origine "arabe" se sont montrés plus sévères lors des évaluations globales que les juges qui avaient une représentation alternative en tête, bien qu'ils n'aient pas trouvé autant de fautes grammaticales. Il est possible que les représentations négatives qui circulent dans la société française envers les personnes "arabes" aient provoqué un "snap judgement"[12] chez certains juges.

2. Application des résultats au 'monde réel'

Les différences observées quant aux erreurs individuelles captées dans chaque groupe de traitement suggèrent que l'évaluation des compétences en langue n'est pas un processus mécanique, et objectif, dans laquelle une erreur a un même poids chez n'importe quel locuteur. En effet, notre perception est sélective, fort influencée par le contexte et les caractéristiques uniques de chaque locuteur. Les erreurs faites au début ou à la fin d'un récit ont tendance à être plus notées, et, certains erreurs ont tendance à ressortir plus chez certains locuteurs. Ces tendances soulignent la complexité des évaluations selon les critères complexes et holistiques comme pour l'examen DELF et la difficulté à

12 Jugement instantané

étudier l'impartialité des juges.

Notre étude ne nous permet pas de dire si les différences entre les erreurs trouvées par les juges se multiplierait d'une manière linéaire lors d'une évaluation plus longue. Autrement dit, si l'enregistrement avait duré dix minutes au lieu d'une seule, est-ce que l'écart entre l'étudiante C et les deux autres aurait continué à s'accroître? Cette question est très pertinente et il serait intéressant de l'aborder lors des études futures.

Notre étude ne nous permet pas non plus de dire si les différences entre les erreurs trouvées par les juges se répéteraient dans une étude menée avec les juges formés en FLE. La formation des juges pour les examens comme le DELF et l'OPI est rigoureuse et peut avoir un impacte sur leur capacité à éviter les stéréotypes.

Une conclusion perturbante de cette étude était l'influence mise en évidence pour le stéréotype 'arabe'. Bien que le clip audio soit exactement le même, les juges avaient tendance à être plus sévères envers l'étudiante A lorsqu'ils connaissaient son origine "arabe". Les différences ne sont pas quantitativement important mais sont statistiquement significatives, ce qui indique que cette origine pourrait effectivement influencer les jugements langagiers et académiques.

Nous voulons réitérer le fait que trouver une telle influence ne dénote pas forcement des préjugés individuels des juges. En effet, cette influence pourrait être causée par des stéréotypes circulant dans la société. Il est possible qu'un meilleur repérage de ces stéréotypes et de leur influence possible permettrait de mieux former les juges et ainsi combattre ces aspects subjectifs.

3. Limites

La méthodologie employée dans cette étude était exploratoire. En réfléchissant sur nos protocoles, nous avons révélé des points forts ainsi que des points faibles. D'un côté, nous avons pu faire de nombreux analyses novatrices, parvenant à des résultats intéressants. De l'autre côté, il existe de

nombreuse limites importantes de cette méthodologie.

1. La durée des clips audio : Chaque clip audio dure de 52 à 55 secondes. Cela ne donne pas aux juges beaucoup de temps pour noter les erreurs et les a probablement empêcher d'en capter plus.

2. Le texte lu par les étudiantes étrangères : Nous avons utilisé un texte écrit, ce qui n'est pas semblable à une énoncée authentique. Il vaut mieux partir d'un texte oral pour les enregistrements.

3. La taille des échantillons : Nous voulions trouver au moins 50 juges pour chaque variante du protocole, mais à cause des variations de l'effectif d'étudiants dans les cours, nous n'avons pas pu équilibrer ce chiffre. Notamment, nous n'avons pas pu atteindre un échantillon de taille suffisant pour le groupe C pour certains analyses. On doit reconnaître que les conclusions dégagées quant aux analyses par groupe de traitement auraient pu être plus fiables avec un échantillon plus important.

4. La méthode expérimentale : les conditions d'évaluation lors des expériences ne ressemblent pas beaucoup à celles du 'vrai monde'. L'extrait de l'étudiante est très court, il n'y a pas la possibilité d'interagir avec elle, et les juges ne sont pas formés. Nous pensons que ces limites sont acceptables étant donné la nature de nos objectifs. Nous ne cherchions pas à trouver un degré précis d'influence, imitant les conditions exactes d'un examen, ni de pointer du doigt à un groupe de juges en particulier. Nous voulions plutôt apporter une analyse expérimentale à une situation où, dans son contexte réel, l'énorme quantité de variables rend très difficile la mise en œuvre d'une comparaison. Notre approche expérimentale nous a permis de cibler les influences de l'accent et de l'origine et de déterminer si ces variables pouvaient, dans un certain contexte, avoir un impact sur les évaluations de compétence langagière. Une prochaine étape serait d'essayer un protocole semblable avec des juges formés en FLE et un clip plus long.

5. Complexité de l'étude : Le fait de devoir analyser trois groupes de traitement et ensuite de diviser un groupe de traitement en deux rendait assez complexes les analyses ainsi que l'explication des résultats.

CONCLUSION

L'évaluation de la compétence langagière est une pratique complexe et exigeante. Les critères actuels pour évaluer les apprenants de langue exigent un jugement professionnel bien formé et réfléchi. Or, il existe des processus inconscients et automatiques qui peuvent influencer les évaluateurs et les empêcher d'être impartiaux. À la suite de ce travail, nous avons réalisé une certaine avancé pour mettre en évidence l'influence de l'origine ethnique imaginée du candidat. Dans un contexte expérimental, nous avons constaté que ce facteur peut exercer une influence sur les évaluations des juges, dans leur repérage des erreurs commises par les apprenants de langue et dans leurs évaluations globales concernant leur compétence langagière et leur potentiel académique.

Enfin, nous estimons que cette question de l'impartialité dans les examens de compétence langagière mérite plus d'études et de discussion. Notre avons à peine effleuré les possibilités des questions de recherche et des variables potentiellement influentes. Nous voyons un fort potentiel dans la méthodologie adaptée du 'Matched Guise Technique' pour pouvoir isoler les variables et cibler l'aspect "imaginé" de notre perception. Ce travail est délicat car nous pouvons nous retrouver confronté à des résultats 'gênants'. Pourtant, ces travaux sont importants car ils nous donnent des pistes de réflexion, qui offre une contribution positive à l'avenir.

BIBLIOGRAPHIE

Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behaviour as predictors of interpersonal consequences : A meta-analysis. *Psychological Bulletin, 111.* 256-274.

Aries, P. (1948). *Histoire des populations françaises*. Paris : Seuil.

Bargh, J. A., & Chartrand, T. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7). 462-479.

Bargh, J. A. & Williams, E. (2006). The automaticity of social life. *Current Directions in Psychological Science, 15*(1). 1-4.

Berthele, R. (2011). The influence of code-mixing and speaker information on perception and assessment of foreign language proficiency: An experimental study. *International Journal of Bilingualism, 16*(4). 453-466.

Campbell-Kibler, K. (2008). I'll be the judge of that : Diversity in social perceptions of (ING). *Language in Society, 37.* 637-659.

Campbell-Kibler, Kathryn. 2010. New directions in sociolinguistic cognition. *University of Pennsylvania Working Papers in Linguistics* 15(2). 31–39.

Campbell-Kibler, K. (2012). The Implicit Association Test and sociolinguistic meaning. *Lingua, 122*(7). 753-763.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Chassard, D. (2006). *Le test des associations implicites (IAT) ou la mesure des évaluations automatiques d'objets d'attitudes : contribution critique à la validité des effets IAT d'attitudes.* Thèse non-publié. Laboratoire de Psychologie des Universités Nancy 2 et Metz Paul-Verlaine : Université Nancy 2. Accessed Aug 19, 2013 at https://docs.google.com/viewer?a=v&pid=gmail&attid=0.1&thid=14086e5db6230dae&mt=application/pdf&url=https://mail.google.com/mail/u/0/?ui%3D2%26ik%3D14811630bc%26view%3Datt%26th%3D14086e5db6230dae%26attid%3D0.1%26disp%3Dsafe%26realattid%3Df_hkfbfdjh0%26zw&sig=AHIEtbRUv1tJFoD6HoFA7bs7MzEt93obhw

Council of Europe. (2001). *DELF Cadre Européen Commun de Reference, Niveau B*2. Accessed December 12, 2012 at http://flehccalbaiulia.wordpress.com/category/enseignement-du-francais/documents-utiles-aux-eleves/

de Damasio, A. (2002). The neural basis of social behavior : Ethical implications. In W. Glannon, *Defining right and wrong in brain science : Essential readings in neuroethics,* pp. 175-178. New York : Dana Press.

Dervin, F. & Salmi, E. (2007). *Évaluer les compétences langagières et interculturelles dans l'enseignement supérieur.* Turku : Finlande. Département d'études françaises à l'Université de Turku.

Desvaux, A. (2005). L'asymptote du français avancé: Les difficultés résistantes. *Études de linguistique appliquée, 138*(2). 223-241.

Fazio, R. & Olson, M. (2003). Implicit measures in social cognition research : Their meaning and use. *Annual Review of Psychology, 54*. 297-327.

Gauthier, E. (2011). *Les fondaments naturels du jugement moral : Rationalisme et sentimentalisme à l'ère des neurosciences*. Master's paper. Montréal : Université du Québec à Montréal.

Gazzaniga, M. (2005). *The Ethical Brain.* New York : Dana Press.

Geeraerts, D., Kristiansen, G, & Peirsman, Y. (2010). *Advances in cognitive sociolinguistics.* New York : de Gruyter.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education, 24*. 355-392.

Greenwald, A. McGhee, D. & Schwartz, J. (1998). Measuring individual differences in implicit cognition : The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(6). 1464-1480.

Gralon. (2012). *Le Panthéon : présentation et histoire.* Accessed September, 2012 at http://www.gralon.net/articles/art-et-culture/architecture/article-le-pantheon---presentation-et-histoire-5915.htm

Le Bon, G. (1895). *Lois psychologiques de l'évolution des peuples*. Paris : Félix Alcan.

Halleck, G. (2008). Interrater reliability of the OPI : Using academic trainee raters. *Foreign Languae Annals, 29*(2). 223-234.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458-484.

Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*(4), 865-892.

Jenkins, S & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of non-verbal, paralinguistic and verbal behaviours in assessment decisions.

The Modern Language Journal, 67. 90-107.

Jones, C. P. (2000). Levels of racism : a theoretical framework and a gardener's tale. *AM J Public Health, 90*(8), 1212-1215.

Jugé, T. S. (2009). *Racism in France : The Civilizing Mission of Whiteness.* Dissertation, Pasadena City College Social Sciences Division. Accessed at http://ssrn.com/abstract=1505644

Kang, O. (2008). Ratings of L2 oral performance in English : Relative impact of rater characteristics and acoustic measures of accentedness. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 6. 181-205.*

Lambert, W. E. (1967). A Social Psychology of Bilingualism. *Journal of Social Issues*, *23*. 91-108.

Ministère de l'interieur, de l'outre-mer, des collectivités territoriales et de l'immigration. (2011). *Circulaire n. NOR IOCN1132114C* : *relative au niveau de connaissance de la langue française requis des postulants à la nationalité française.* Ministère de l'intérieur, Paris, France. Accédé le 3 juillet, 2013, http://circulaire.legifrance.gouv.fr/pdf/2011/12/cir_34304.pdf

McConnell, A. et Leibold, J. (2001). Relations among the Implicit Association Test, discriminatory behaviour, and explicit measures of racial attitudes. *Journal of experimental social psychology, 37.* 435-442.

McGurk, H. & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264.* 746-748.

McNamara, T. (1996). *Measuring second language performance.* London: Longman

McNamara, T. (2001). Language assessment as social practice : Challenges for research. *Language Testing, 18*(4). 333-349.

Nguyen, B. (1993). Accent discrimination and the test of spoken english : A call for an objective assessment of the comprehensibility of nonnative speakers. *California Law Review (81)*5. 1325-1361.

Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18. 62–85.

Omi, M. & Winant, H. (1994). *Racial Formation in the United States : From the 1960s to the 1980s.* New York : Routledge.

Rubin, D. (1992). Nonlanguage factors affecting undergraduates' judgements of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4). 511-531.

Shiffrin, R., Dumais, S., & Schneider, W. (1981). Characteristics of automatism. *Attention and*

Performance, 223. 238-251.

Shohomy, E. (2001) *The power of tests. A critical perspective on the uses of language tests.* Singapore : Longman.

Silberman, R., Alba, R., & Fournier, I. (2006). Segmented assimilation in France? Discrimination in the labour market against the second generation. *Ethnic and Racial Studies, 30*(1), 1-27.

Simon, P., & Clément, M. (2006). Comment décrire la diversité des origines en France? Une enquête exploratoire sur les perceptions des salariés et des étudiants. *Population et Sociétés*, *425*, 1-4.

Simon, P. (2008). Nationalité et origine dans la statistique française. Les catégories ambigues. *Population, 53*(3), 541-567.

Strand, E. A. (1999). Uncovering the rôle of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*(1), 86-99.

Surface, E. & Dierdorff, E. (2003). Reliability and the ACTFL Oral Proficiency Interview : Reporting Indices of Interrater Consistency and Agreement for 19 Languages. *Foreign Language Annals, 36*(4), 507-519.

Sutherland, G. (1996). Assessment : Some historical perspectives. In H. Goldstein & T. Lewis (Eds.), *Assessment : Problems, developments and statistical issues.* Chichester, England : Wiley.

Thomas, E. (2002). Sociophonetic applications of speech perception experiments. *American Speech, 77*(2). 115-147.

Thompson, I. (2008). A study of the interrater reliability of the ACTFL oral proficiency interview in five european languages : Data from ESL, French, German, Russian, and Spanish. *Foreign Language Annals, 28*(3), 407-422.

Von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. *Psychology, 27.* 177-254.

Von Hippel, W., Silver, L., Lynch, M. (2000). Stereotyping against your will : The role of inhibitory ability in stereotyping and prejudice among the elderly. *Personality and Social Psychology Bulletin, 26*(5), 523-532.

Winke, P. Gass, M. & Myford, C. 2011. *The relationship between raters' prior language study and the evaluation of foreign language speech samples*. Princeton, NJ : Educational Testing Service.

ANNEXES

Appendice A: Questionnaire pour le présélection des étudiantes
(comme c'était un questionnaire en ligne, ci-dessous le texte seulement)

Sept étudiantes étrangères ont lu un extrait d'un texte. Vous allez les évaluer sur leur prononciation.

D'abord, lisez attentivement le questionnaire ci-dessous.

Ensuite, écoutez deux fois le clip audio que je vous ai envoyé - clip qui contient l'enregistrement des sept étudiantes.

1ère fois : Écoutez-le du début jusqu'à la fin sans rien noter.

2ème fois : Reprenez l'écoute, et cette fois-ci, faites une pause entre chaque étudiante pour indiquer votre réponse.

* Intelligibilité parfaite, compréhension aussi facile que si j'écoutais un français
* Très bonne intelligibilité, facile à comprendre
* Moyenne, je peux comprendre tout mais avec un peu d'effort
* Peu intelligible, assez difficile à comprendre

Appendice B : Le Texte à être lu par les étudiantes étrangères pour la présélection

Le **Panthéon** est l'un des monuments les plus célèbres de Paris. Il est situé sur la montagne Sainte-Geneviève dans le 5e arrondissement.

S'inspirant du Panthéon d'Agrippa à Rome, ce monument a un dôme majestueux et trois coupoles. Il a une hauteur de 83 mètres. N'ayant pas beaucoup de fenêtres, il est donc froid et sombre à l'intérieur.

Depuis 1885, il est devenu un temple laïc destiné à honorer les français célèbres et à rappeler des événements marquants de l'histoire de France. À l'intérieur, il y a une crypte, mais aussi des tableaux et des fresques illustrant la vie de sainte Geneviève.

Appendice C : Questionnaire (avec carte)

Expérience : apprenants de français à Grenoble
Partie 1 : erreurs grammaticales
Consignes: Vous allez écouter un clip audio d'une étudiante étrangère qui apprend le français. Nous faisons une expérience pour savoir si l'évaluation de son français est la même quand elle est faite par des profs de français ou par d'autres personnes.

Vous n'allez entendre le clip audio qu'une seule fois. Pendant que vous écoutez, notez les erreurs grammaticales que vous repérez dans le tableau ci-dessous.

Attention:

- On cherche des erreurs grammaticales (faute de conjugaison de verbe, genre, temps, etc.). À ne pas confondre avec des erreurs de prononciation.
- Il n'y a pas forcement 10 erreurs
- Nous vous remercions d'écrire de manière lisible

# Faute	**Description,** ex. *J'ai allé*
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Partie 2 : Évaluation globale

1. Cette étudiante étrangère démontre une maîtrise de la langue qui lui permet de bien s'exprimer.

Tout à fait d'accord	**D'accord**	**Pas d'accord**	**Pas du tout d'accord**
☐	☐	☐	☐

2. Malgré son accent, cette étudiante étrangère possède une prononciation intelligible qui n'entraîne pas de problèmes de compréhension.

Tout à fait d'accord	**D'accord**	**Pas d'accord**	**Pas du tout d'accord**
☐	☐	☐	☐

3. Selon sa performance dans le clip audio, cette étudiante étrangère serait capable de suivre un cursus universitaire en France.

Tout à fait d'accord	**D'accord**	**Pas d'accord**	**Pas du tout d'accord**
☐	☐	☐	☐

Partie 3 : Origine de l'étudiante

2. Selon vous, de quel pays ou éventuellement de quelle zone géographique vient cette étudiante ? Indiquez-le par une croix sur la carte ci-dessous.

2) comment appelleriez-vous ce pays/cette zone géographique ?_____

Partie 4 : votre profil

***** Date de naissance _____ * Pays de naissance : _____Langue(s) maternelle(s) ____

* Langue(s) parlé(s) par vos parents : Père_____Mère_____

* Parlez-vous une autre langue que le français ? Laquelle ? Dans quelles circonstances ? :

_____

* Avez-vous vécu plus de trois mois dans un autre pays que la France ?

- Si oui, de quel(s) pays s'agit-il ? _____

Appendice D : Évaluation Orale Delf B2

Monologue suivi : Poser un problème

Peut dégager la problématique générale posée par le document	0	0,5	1	1,5	2		
Peut développer la problématique : dégager ses différents aspects, les illustrer avec des exemples	0	0,5	1	1,5	2	2,5	3
Peut structurer sa présentation de façon à mettre en évidence la progression de la pensée	0	0,5	1	1,5	2	2,5	3

A noter : Le texte proposé n'est qu'un document déclencheur. Vous pouvez développer librement la problématique, introduire idées, arguments, exemples suggérés par le document sans toutefois exprimer de manière explicite son opinion personnelle.

Interaction : Présenter son opinion, la défendre si nécessaire

Peut exprimer ses idées de façon précise, nuancée. Peut apporter des précisions si nécessaire.	0	0,5	1	1,5	2	2,5	3
Peut réagir aux propos d'autrui pour montrer son accord/désaccord (catégorique ou nuancée) Peut défendre sa position (la maintenir, la modifier, la nuancer)	0	0,5	1	1,5	2	2,5	3

Evaluation de la compétence langagière (pour l'ensemble de l'épreuve)

Maîtrise des structures de la langue Le candidat est capable d'utiliser les structures de la langue nécessaires à l'expression d'une pensée nuancée malgré des hésitations et des fautes n'entraînant pas de problèmes de compréhension de son message	0	0,5	1	1,5	2	2,5	3	3,5	4
Maîtrise du lexique Peut utiliser une variété large de vocabulaire en respectant les registres de langue. Est capable, lorsqu'un mot lui manque de reformuler sa pensée de manière efficace.	0	0,5	1	1,5	2	2,5	3	3,5	4
Maîtrise de la prononciation et de l'intonation Possède, malgré son accent, une prononciation intelligible qui n'entraîne aucun problème de compréhension. Est capable de rendre son discours plus expressif par une bonne maîtrise de l'intonation.	0	0,5	1	1,5	2	2,5	3		

Appendice E : À lire avant la passation du protocole

Bonjour,
Je fais de la recherche sur les apprenants de français à Grenoble.

M/Mme _____ a gentiment accepté que je viens pour faire passer un questionnaire qui porte sur les compétences en français d'une fille qui apprend le français.

- Personne n'a jamais fait cette expérience ou connais la fille qui va parler ? (voir photo)
-

OK, Lisons les consignes sur le questionnaire.
Partie 1 : erreurs grammaticales
Consignes: Vous allez écouter un clip audio d'une étudiante étrangère (*dont la langue maternelle est arabe/chinoise/portugaise)* qui apprend le français.

Nous faisons une expérience pour savoir si l'évaluation de son français est la même quand elle est faite par des profs de français ou par d'autres personnes.

Vous n'allez entendre le clip audio qu'une seule fois.

Pendant que vous écoutez, notez les erreurs grammaticales que vous repérez dans le tableau ci-dessous.

Attention:

- On cherche des erreurs grammaticales (faute de conjugaison de verbe, genre, temps, etc.). À ne pas confondre avec des erreurs de prononciation.
- Il n'y a pas forcement 10 erreurs
- Nous vous remercions d'écrire de manière lisible

Après avoir noté vos réponses, vous pouvez tournez la page et remplir le verso. Vous aurez suffisamment du temps pour le finir.

(faire passer le clip)
(donne 7 minutes)
(ramasser les questionnaires)

Merci bien pour votre participation.

Appendice F : Catégorisation des pays/zones géographiques

zones "asiatiques"	Pays/zones en Asie
zones "blanches"	Pays/zones en Amérique du Nord, Amérique du Sud, Europe, Austrasie
zones "arabes"	Pays/zones basés sur le "member states of the arab league"[13] : Algérie, Bahreïn, Comores, Djibouti, Égypte, Irak, Jordan, Koweït, Liban, Libye, Mauritanie, Maroc, Oman, Palestine, Qatar, Arabie Saoudite, Somalie, Soudan, Syrie, Tunisie, UAE, Yémen
zones "noires"	Tous les pays en Afrique pas codés comme zone/pays arabe

13 en.wikipedia.org/wii/Arab-world

Résumé

Cette étude porte sur la fiabilité de l'évaluation de la compétence orale en français langue étrangère. Nous voulions savoir si l'origine perçue d'un locuteur apprenant le français, qu'il s'agisse de l'origine réelle ou imaginaire, influençait les évaluations sur sa compétence orale portées par un échantillon de juges français. L'approche adoptée est une adaptation de la technique du "matched guise" (Lambert, 1967). Notre démarche comporte deux étapes. D'abord, nous demandons à trois grands groupes de juges francophones (des étudiants de licence des universités grenobloises, au total N =343) d'évaluer la production orale de trois locutrices non-natives, d'origines nationales différentes (chinoise, japonaise, syrienne), lisant un même texte comportant les mêmes erreurs grammaticales. Ensuite, nous nous centrons sur une seule des locutrices non-natives, d'origine syrienne, et nous comparons les jugements des évaluateurs qui connaissaient son origine "arabe" et ceux qui ne la connaissaient pas, isolant ce facteur comme variable indépendante. Ces deux axes nous permettent d'aborder la notion d'origine dans ses dimensions "réelles" (trois locutrices non-natives d'origines différentes) et "imaginées" (des perceptions différentes de l'origine un seul locuteur).

Les analyses statistiques démontrent que parmi les juges écoutant l'enregistrement de l'étudiante syrienne, ceux qui perçoivent son origine comme "arabe" lui attribuent des évaluations plus sévères que ceux qui ne la perçoivent pas comme "arabe", bien que par ailleurs ils décèlent moins de fautes dans sa production orale. Cette double tendance suggère que les jugements de non spécialistes sur les compétences langagières en langue étrangère peuvent être biaisés par des stéréotypes et des préjugés.

Mots-Clés

évaluation des compétences en langue étrangère
impartialité des jugements
automaticité
perception de la parole
cognition sociolinguistique
stéréotypes
préjugés