

Constitution d'un Corpus de Français Langue Etrangère destiné aux Apprenants Allemands

Camille FAUTH¹, Anne BONNEAU¹, Odile MELLA¹, Vincent COLOTTE¹, Dominique FOHR¹, Denis JOUVET, Yves LAPRIE¹, Jürgen TROUVAIN²

¹ Groupe Parole, LORIA CNRS UMR 7503, Nancy, France

² Université de la Sarre, Sarrebrück– Allemagne

camille.fauth@gmail.com

bonneau@loria.fr

Introduction

Il est clair que la langue maternelle (maintenant L1) influence l'apprentissage d'une seconde langue (maintenant L2) et ce à tous les niveaux linguistiques tels que le lexique, la morphosyntaxe, la pragmatique mais aussi et particulièrement la phonétique (par ex. Flege & Davidian 1984, Flege 1995). Il est presque impossible pour un apprenant d'atteindre un niveau qui soit comparable à celui d'un locuteur natif. Ceci est vrai pour tous les niveaux linguistiques précédemment cités et l'est peut-être encore plus pour ce qui concerne les aspects phonétiques et phonologiques. La prononciation d'une L2 est influencée par le système phonologique et par les connaissances phonétiques de la L1 (Best, 1995 ; Barry et al. 2007 ; Dupoux et al. 2008). Ces influences peuvent conduire à des interférences dans la production et la perception de sons, de mots ou de phrases en L2. Presque tous les apprenants d'une langue étrangère, parlant avec un « accent étranger », sont confrontés à une réduction de leur expressivité verbale (Hirschfeld et Trouvain, 2007). Cette impression d'« accent étranger » est déterminée par le nombre d'erreurs et leur ampleur. Divers facteurs externes sont responsables du degré d'accent étranger, l'âge d'apprentissage de la L2 étant l'un des plus prépondérants (Piske et al., 2001). Bien que le degré d'accent étranger soit généralement corrélé avec une intelligibilité réduite, un fort accent étranger ne conduit pas inévitablement à une intelligibilité réduite (Munro & Derwing, 1995; Wittemann et al, 2011). Les individus se comportent très différemment et pas toujours en fonction de leur niveau de compétence phonétique (Mehlhorn, 2005) et leur talent phonétique (Jilka, 2009). L'un des problèmes presque universel pour les locuteurs non natifs est que leur discours est moins fluide en L2 que dans leur L1. Celui-ci est marqué par une vitesse d'élocution plus lente, des pauses plus importantes et plus longues et des disfluences qui peuvent être notamment des hésitations sur le son à produire ou des faux départs (Cucchiari et al., 2000). Le niveau de compétence dans une langue est hautement corrélé avec la vitesse d'élocution (Gut, 2009). La prosodie est sans doute un trait très saillant pour les locuteurs natifs qui entendent de la parole non native (Jitka, 2007). Le phénomène de « surdité accentuelle », par exemple, (Dupoux et al., 2001) décrit pour des locuteurs ayant le français en L1 conduit à des problèmes considérables en production et en perception de la parole non native.

La plupart des corpus en langue se focalisent sur les phénomènes linguistiques écrits et concernent l'anglais (voir le site web : "Learner corpora around the world" de l'Université de Louvain - Belgique). La recherche phonétique sur l'acquisition d'une L2 est généralement orientée vers l'étude des phénomènes segmentaux et la plupart des études ont également l'anglais comme langue cible. Les modèles de parole en L2 actuels - voir par exemple Speech Learning Model (Flege, 1995) ou Best's Perceptual Assimilation Model (Best, 1995) – négligent bien souvent les aspects prosodiques. Notre étude concerne le français en tant que langue seconde et s'inscrit dans un projet plus vaste mené en partenariat avec une université allemande, dont l'un des buts est le développement de l'apprentissage des langues par ordinateur. (Projet ANR-DFG – Agence Nationale de la Recherche et Deutsche Forschungsgemeinschaft attribué à l'équipe Parole du LORIA UMR 7503, Nancy – France et à l'Equipe de Linguistique Computationnelle et de Phonétique FR 4.7 de l'Université de la Sarre Sarrebruck – Allemagne) dans lequel le français et l'allemand sont des langues cibles. Pour la paire allemand-français, peu de corpus parallèles sont disponibles. Nous présentons ici l'élaboration d'un corpus de productions orales de locuteurs natifs et non natifs pour la paire allemand-français. Notre corpus entend mettre au jour les

déviations phonétiques et phonologiques que les locuteurs allemands produisent lorsqu'ils apprennent le français. Ce projet entend étudier les difficultés que les locuteurs français rencontrent lorsqu'ils apprennent l'allemand, et réciproquement. Aussi, cinquante locuteurs allemands seront recrutés dans des milieux universitaires et scolaires (niveau lycée) en Allemagne et cinquante locuteurs français dans les mêmes milieux en France. Il s'agit pour les deux populations de produire d'une part le corpus en langue étrangère (en langue française pour les locuteurs allemands et en langue allemande pour les locuteurs français) mais également le corpus en langue maternelle (en allemand pour les allemands et en français pour les français). Les corpus ainsi obtenus devraient nous permettre d'identifier les difficultés que les locuteurs allemands ou français rencontrent lorsqu'ils apprennent le français ou l'allemand. Les données de contrôle sont doubles puisque l'on pourra à la fois se référer aux productions des apprenants dans leur langue maternelle (ici l'allemand), mais également à celles de locuteurs natifs (ici germanophones). Nous ne présenterons ici que la constitution du corpus en français.

Le corpus complet est divisé en deux sous-corpus (Corpus 1 et Corpus 2) qui sont relativement similaires en termes de matériel phonétique. Le Corpus 1 a été enregistré pour tester nos hypothèses en ce qui concerne les interférences entre la L1 et la L2 des locuteurs pour vérifier que la longueur de notre corpus était adaptée aux locuteurs, pour tester le protocole d'enregistrement et enfin évaluer les performances du système d'alignement automatique dans le cas du français prononcé par des locuteurs allemands. D'un point de vue plus technique, ce premier corpus a également été l'occasion de tester le matériel d'enregistrement. Les premiers résultats obtenus ont permis d'élaborer finement le Corpus 2.

La constitution d'un tel corpus répond à trois motivations. (i) Les corpus sont constitués à partir de connaissances phonétiques fines afin de prévoir les possibles erreurs faites et/ou difficultés rencontrées par les apprenants et dans le but d'étudier plus particulièrement les erreurs phonétiques que produisent les locuteurs allemands en français langue étrangère (ii) Les corpus pourront ensuite être utilisés par la communauté scientifique pour la reconnaissance automatique de la parole non-native, ce qui est, on le sait, particulièrement difficile (voir par exemple Goronzy et al. 2001, van Doremalen et al. 2009, Bouselmi et al. 2011). (iii) Nous mettrons à disposition de la communauté scientifique deux corpus oraux symétriques et annotés de productions en français (natifs et non natifs).

Enfin, ce corpus devrait également permettre de tester les systèmes d'alignement automatique texte/audio sur de la parole non native. Les prononciations possibles à aligner ont été obtenues à partir de la parole native, le système d'alignement devrait donc rencontrer des difficultés pour aligner automatiquement de la parole non native dans la mesure où des prononciations non canoniques y apparaissent plus fréquemment. Ce corpus, grâce à un nouvel inventaire des prononciations déviantes produites par des locuteurs non natifs, devrait donc permettre à terme d'améliorer le système d'alignement automatique sur de la parole non native. Rappelons que la détection et la correction des erreurs (qui font partie intégrante de l'apprentissage des langues assisté par ordinateur, cadre dans lequel s'inscrit notre projet) s'appuient sur la reconnaissance automatique de la parole ou sur les méthodes de classification (voir par exemple Witt & Young, 2000; Strik et al., 2009).

1 Constitution du Corpus 1

Il est clair que les apprenants d'une langue étrangère peuvent produire de nombreuses déviations phonétiques et phonologiques, par exemple l'insertion d'un coup de glotte, l'aspiration des occlusives ou l'usage incorrect de la liaison (mais voir aussi la littérature citée précédemment en introduction). Nous avons choisi de constituer un premier corpus (à partir de maintenant Corpus 1) pour sélectionner ensuite plus finement les différents phénomènes cibles qui pourront faire l'objet d'une étude plus systématique (à partir de maintenant Corpus 2). Ces phénomènes ont été sélectionnés sur la base du degré attendu de déviation par rapport à la norme (ou à la réalisation des natifs) et sur leur fréquence et ce au niveau segmental et supra-segmental.

Avant les enregistrements, les quatorze locuteurs ont répondu à un questionnaire précisant leur biographie linguistique (L1, nombre de L2, durée d'apprentissage et diplômes ou certificats dans chaque L2) ainsi qu'une auto-évaluation (motivation pour l'apprentissage des langues, niveau...). Tous les

locuteurs ont réalisé quatre tâches. Pour la première tâche, il était demandé au locuteur de lire à voix haute une phrase. Vingt-cinq phrases différentes lui étaient proposées. Lors de la deuxième tâche, le locuteur entendait une phrase prononcée par un locuteur natif. Il lui était ensuite demandé de répéter (la phrase était également inscrite sur l'écran) cette même phrase à voix haute et ceci pour vingt-cinq phrases différentes. La tâche de lecture a été proposée pour saisir «à vif» la prononciation de l'apprenant mais également pour observer les différents processus prosodiques. La tâche de répétition a été pensée non seulement pour s'affranchir des erreurs dues à la graphie mais également pour se rapprocher des exercices construits par les logiciels d'apprentissage des langues par ordinateur dans lesquels l'utilisateur est généralement amené à répéter des phrases. Enfin, les deux modalités ont été retenues afin d'observer si l'apprenant peut améliorer sa prononciation en langue étrangère s'il entend un locuteur natif prononcer au préalable la phrase qu'il aura à produire.

Les phrases ont été conçues pour couvrir : (i) tous les phonèmes du français dans les contextes pertinents pour permettre un inventaire phonémique propre à chaque locuteur, (ii) les phénomènes phonétiques segmentaux et supra-segmentaux les plus importants du français langue étrangère. Certains de ces phénomènes seront observés à l'aide de paires minimales.

La troisième tâche était une mise en condition de réalisation de focus, contrastif ou non contrastif, à l'aide d'une phrase à prononcer en réponse à une question posée par un locuteur natif. La partie de la phrase qui était supposée être mise en relief était en plus écrite en lettres majuscules. Six phrases étaient présentées à partir desquelles différentes parties étaient mises en relief et ceci pour tester la compétence de nos apprenants à utiliser ou non le schéma prosodique adéquat.

La quatrième et dernière tâche consistait à lire à voix haute deux textes : *les trois petits cochons* et un texte plus technique d'environ dix phrases et ceci notamment pour étudier la prosodie, la fluence et la vitesse d'élocution.

1.1 Sujets et enregistrements

Pour le Corpus 1, nous avons enregistré sept locuteurs allemands apprenant le français en respectant deux classes d'âge (voir Table 1). Notre échantillon de locuteurs comprend des apprenants débutants (niveau A2 en fonction du CEFRL ; voir leur site web) et des apprenants avancés (niveau C2). Parmi les débutants, nous avons retenu des adolescents avec deux ou trois années d'enseignement de la L2 en milieu scolaire et des étudiants ayant bénéficié d'un enseignement de la L2 à l'université. Les locuteurs avancés étaient tous des étudiants. Un consentement éclairé a été obtenu auprès des sujets (et des parents pour les adolescents), nous permettant d'utiliser les données à des fins scientifiques et garantissant l'anonymat des participants. Afin de comparer les productions des apprenants avec des productions de locuteurs natifs, nous avons également enregistré sept locuteurs francophones de classe d'âge et de niveau scolaire équivalent. Les deux cohortes de locuteurs ont produit le même corpus.

Les locuteurs produisent systématiquement le corpus en L2 mais également en L1.

Table 1: Groupes de sujets qui ont été enregistrés pour le Corpus 1 en fonction de leur L1 et de leur L2 (F=Français, A=Allemand), de leur niveau en L2 et de leur âge.

# sujets	L1	âge	# sujets	L1	L2	Niveau	âge
5	F	18-30 ans	3	A	F	Débutants	18-30 ans
			2	A	F	Avancés	
2	F	15-16 ans	2	A	F	Débutants	15-16 ans

Les enregistrements, de haute qualité, ont été effectués à l'aide du logiciel Corpus Recorder (voir le site web de l'équipe Parole du LORIA). Ce logiciel est dédié à l'enregistrement de listes de phrases. Il permet un paramétrage précis du protocole d'acquisition (ordre aléatoire, écoute ou non de sons d'exemples...). Il

intègre notamment un contrôle du gain tout au long de la session d'enregistrement. En sortie, il génère une liste de fichiers au format Wav, automatiquement renommés en fonction de la phrase produite et avec les informations (âge, sexe, nationalité...) relatives au locuteur. Ce dernier point facilite le stockage et l'exploitation des fichiers.

Le sujet était assis dans une pièce calme et avait pour consigne de produire les phrases à une intensité confortable. La collecte des données a été réalisée à l'aide d'un microphone-casque (AKG C520) et une Audiobox (M Audio Fast track) reliés à un ordinateur portable Microsoft. Les sessions d'enregistrement ont été soigneusement contrôlées pour une qualité constante. Si nécessaire, le gain était ajusté pour éviter que le signal ne soit saturé ou trop faible comme cela peut être le cas lorsque les sessions d'acquisition de données sont longues. Après avoir été enregistrés les différents fichiers audio du Corpus 1 ont été alignés automatiquement avec leur transcription orthographique puis corrigés manuellement.

1.2 Alignement texte/audio automatique du corpus

(1) Description de notre système automatique d'alignement texte/audio

Afin de réaliser l'alignement d'un texte sur un signal audio, il faut d'abord générer les différentes prononciations possibles correspondant à ce texte. Cette étape utilise un lexique fournissant pour chaque mot différentes prononciations et un ensemble de règles pour prévoir notamment les liaisons ou les pauses éventuelles que le locuteur pourrait faire. Lors de la phase suivante, un alignement forcé est réalisé en utilisant des techniques issues de la reconnaissance automatique de la parole afin de déterminer la suite de phonèmes qui correspond le mieux au signal audio. Ces techniques sont principalement fondées sur une modélisation stochastique acoustique des phonèmes par des Modèles de Markov Cachés (en anglais : Hidden Markov Models ou HMMs) et sur l'algorithme de Viterbi. Chaque phonème est représenté par un modèle acoustique à 3 états où chaque état est un mélange de gaussiennes : le premier état modélise le début du phonème, le deuxième la partie centrale du phonème et le dernier état la partie finale. L'algorithme de Viterbi réalise simultanément la détermination de la meilleure séquence de modèles correspondant au signal et la détermination des frontières temporelles de ceux-ci. Le signal audio n'est pas utilisé directement mais prétraité pour calculer tous les centièmes de seconde un vecteur de paramètres spectraux. Les modèles HMMs sont appris sur un grand corpus de parole (plusieurs centaines d'heures d'audio) comprenant un très grand nombre de locuteurs (pour assurer une grande variabilité).

L'alignement forcé est confronté au problème suivant. Pour trouver la bonne étiquette phonétique associée à un segment de signal audio, il faut utiliser des modèles acoustiques précis donc comportant beaucoup de gaussiennes. Malheureusement des études ont montré que pour obtenir des frontières précises, il fallait utiliser des modèles avec peu de gaussiennes (Toledano & Gomez, 2003). Pour pallier ce problème, nous effectuons deux alignements successifs avec deux ensembles de modèles acoustiques. Un premier alignement avec des modèles précis va permettre de sélectionner la meilleure séquence de phonèmes correspondant au signal audio. Puis, cette meilleure séquence sera alignée avec le même signal en utilisant des modèles avec peu de gaussiennes afin d'obtenir les frontières des phonèmes.

Concrètement, notre système d'alignement utilise 36 modèles acoustiques de phonèmes complétés par un modèle de pause, un modèle de respiration et un modèle de bruit (Fohr & Mella 2012). Ces modèles ont été appris sur 223 heures d'audio issues de corpus transcrits provenant d'émission de radio et de télévision construits dans le cadre des projets ESTER (Galliano et al., 2006) et ETAPE (Gravier et al., 2012). Les vecteurs de paramètres acoustiques sont constitués de 39 coefficients MFCC (Mel Frequency Cepstral Coefficients), les modèles précis comportent 32 gaussiennes par état et les modèles permettant de déterminer les frontières, 2 gaussiennes par état.

(2) Difficultés de l'alignement texte/audio dans le cas de la parole non native

Un alignement automatique n'est jamais parfait mais dans le cas de la parole non native des erreurs supplémentaires peuvent survenir. En effet, les prononciations potentielles à aligner ont été générées pour de la parole native. Or le locuteur non natif peut prononcer un nombre de phonèmes différent de celui attendu, par exemple en prononçant une consonne muette («*petit*» prononcé /pətit/) ou en élidant un

phonème non élidable («*belote*» prononcé /blot/). De plus, les locuteurs non natifs ne réalisant pas correctement certains phonèmes du français, comme par exemple les voyelles nasales, leurs réalisations ne correspondent pas aux modèles acoustiques appris. Par ailleurs, les modèles acoustiques ont été appris sur de la parole prononcée par des locuteurs français natifs donc ayant un rythme d'élocution plus rapide que celui d'un locuteur non natif.

Il a donc été nécessaire de faire une correction manuelle de l'alignement automatique du Corpus 1 avant de l'analyser. La totalité du corpus a été alignée en utilisant l'alphabet phonétique SAMPA, (voir site web de l'UCL) puisque cet ensemble de symboles présente l'intérêt principal d'utiliser des caractères utilisables facilement sur ordinateur. Le traitement du signal a été effectué à l'aide du logiciel PRAAT (voir site web de Praat) à partir duquel il a été possible d'ajouter différentes lignes d'informations (ou tier) pour un alignement au niveau de la phrase, du mot et du phonème (voir Figure 1). Dans un premier temps, l'annotateur est chargé de vérifier et de déplacer, à l'aide du signal acoustique et de son spectrogramme associé, les frontières mises en place par l'alignement automatique. Dans un second temps, grâce à une série de diacritiques adaptée du corpus Kiel (voir le site web du corpus Kiel) l'annotateur est amené à noter les différentes variations produites par le locuteur. Cette liste de diacritiques pourra être complétée au fur et à mesure de l'avancement du projet. L'annotateur dispose également d'une ligne *commentaire* qu'il peut utiliser pour ajouter des remarques, cette ligne est toutefois à utiliser avec parcimonie mais peut se révéler utile pour les remarques qui concernent la totalité de la phrase (au niveau de la prosodie notamment).

Table 2: Liste des diacritiques qui permettent la correction et l'annotation manuelle du corpus.

Phénomènes	Diacritiques	Phénomènes	Diacritiques
Omission	X-	Pauses	-#
Insertion	-X	Voisement	_V
Substitution	X-Y	Dévoisement	_0
Hésitation	&mot	Bruits	!
Frontières floues	%	Bruits du locuteur	§
Mots tronqués	/mot	Relâchement de l'occlusion	X_
Glottalisation	q	Accent	‘
Coup de glotte	Q	Relâchement du coup de glotte	Q_
Prosodie	p		

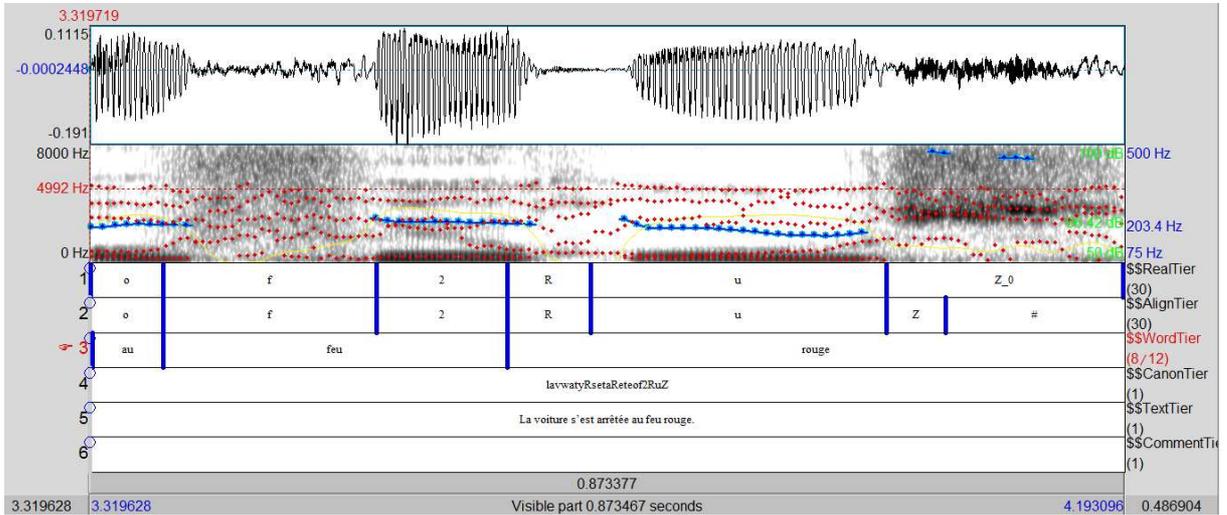


Figure 1: Phrase française lue et produite par un locuteur allemand. « La voiture s'est arrêtée au feu rouge » ou «lavvatyRsetaReteof2RuZ» RealTier : Corrections des frontières de l'alignement automatique et annotations manuelles à l'aide de diacritique ; AlignTier : Alignement automatique au niveau du phonème transcrit en SAMPA ; WordTier : Alignement automatique au niveau du mot ; TextTier : Alignement automatique au niveau de la phrase ; CommentTier : Commentaires éventuels de la part de l'annotateur manuel.

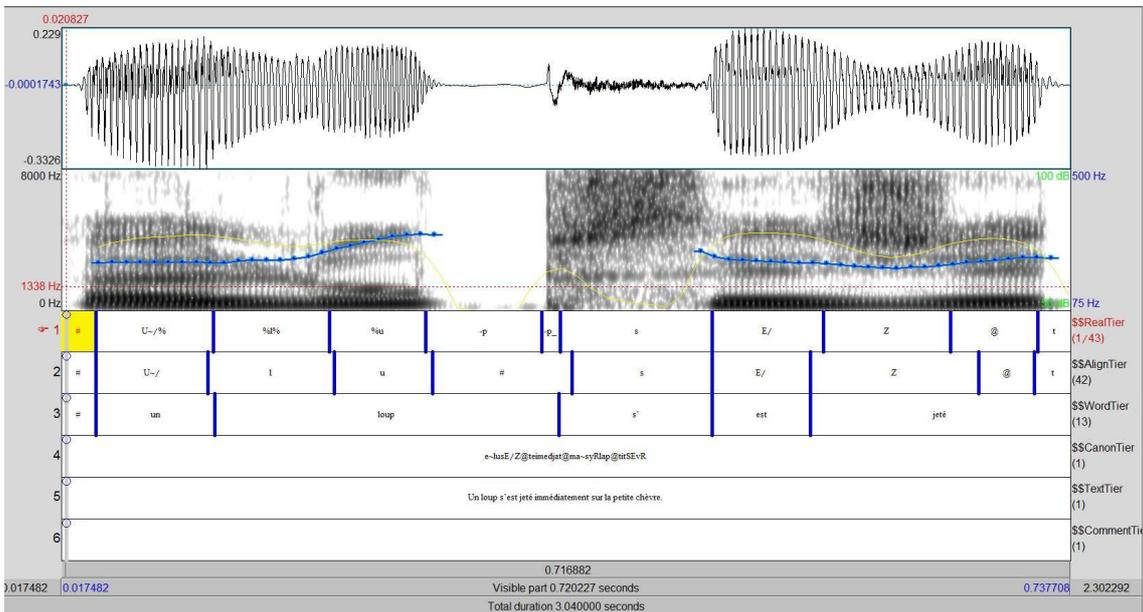


Figure 2: Phrase française lue et produite par un locuteur allemand. « Un loup s'est jeté immédiatement sur la petite chèvre » ou «e~luE/Z@teimedjatma~syRlap@titSEvR» RealTier : Corrections des frontières de l'alignement automatique et annotations manuelles à l'aide de diacritique ; AlignTier : Alignement automatique au niveau du phonème transcrit en SAMPA ; WordTier : Alignement automatique au niveau du mot ; TextTier : Alignement automatique au niveau de la phrase ; CommentTier : Commentaires éventuels de la part de l'annotateur manuel.

1.3 Résultats préliminaires et conclusions à partir du Corpus 1

Le but du Corpus 1 était de vérifier la pertinence d'un phénomène phonétique particulier et éventuellement de relever des phénomènes et/ou déviations non attendus dans les productions des apprenants. Le Corpus 1 a donc volontairement été construit comme un « balayage » parfois un peu large de toutes les difficultés qu'un locuteur allemand pourrait rencontrer en français L2. Nos données suggèrent, comme attendu, que de nombreux phénomènes se sont produits et sont essentiels pour l'amélioration de la prononciation des apprenants en L2.

En ce qui concerne l'étude de la fréquence fondamentale (hauteur et variation), l'analyse du Corpus 1 suggère qu'elle est réduite en L2. Ceci est vrai quel que soit le niveau de compétence en L2. Des données plus précises sont actuellement en cours de traitement et d'analyse (Fauth et Bonneau, 2014 ; Zimmerer et al., 2014).

De plus, nous avons étudié des aspects segmentaux comme le dévoisement des obstruantes en position finale dans la syllabe. Cette étude nous a notamment permis de voir qu'en fonction de la condition de lecture ou de répétition les locuteurs n'adoptent pas la même stratégie. En effet, si en lecture, les apprenants, notamment les débutants dévoisent quasi-systématiquement leurs obstruantes en position finale, en répétition, ils ajoutent un schwa final plaçant ainsi l'obstruante dans une position non finale artificielle. L'obstruante est alors voisée.

L'un des objectifs du Corpus 1 était de tester le protocole d'acquisition du corpus et nous avons pu vérifier que l'orthographe influence la production des apprenants. Pour exemple, la consonne muette du mot *loup* est très souvent prononcée par les apprenants débutants. Ceci renforce donc encore un peu plus la nécessité de conserver une tâche de lecture et une tâche de répétition afin d'étudier différents phénomènes et parfois pouvoir s'affranchir de l'orthographe ou de la méconnaissance du vocabulaire qui peuvent poser des problèmes à nos apprenants et donc provoquer des erreurs de prononciations qui ne sont pas liées à leurs compétences phonétiques en L2.

Enfin, sans surprise, les apprenants produisent plus de faux départs, d'hésitations et de pauses, parfois mal à propos en L2. Leur vitesse d'élocution est également réduite.

A partir de ces premières constatations, nous avons pu déduire des aspects importants, décrits ci-dessus, dans l'interférence L1/L2 et améliorer les contextes dans lesquels apparaissent les phénomènes dans nos données. Nous avons donc modifié le Corpus 1 pour créer un Corpus 2 que l'on souhaite plus performant. C'est de ce corpus dont il est question dans la section suivante.

2 Corpus 2

2.1 Constitution du Corpus 2

La motivation pour le Corpus 2 est la même que celle qui a permis la constitution du Corpus 1, il s'agit d'obtenir de locuteurs allemands des productions en français L2. Des locuteurs francophones natifs seront également enregistrés et ceci toujours dans l'optique de pouvoir comparer finement les productions des apprenants avec celles de locuteurs natifs. Rappelons que ce projet est franco-allemand c'est pourquoi les locuteurs sont systématiquement enregistrés dans leur L1 (français ou allemand) et dans leur L2 (allemand ou français). Les productions des apprenants en L1 (i.e. langue allemande) pourront éventuellement nous servir à affiner nos remarques en nous reportant aux stratégies individuelles qu'ils utilisent dans leur langue maternelle.

Si pour le Corpus 1, notre démarche se voulait volontairement plus généraliste et entendait couvrir toutes les difficultés qu'un apprenant allemand pouvait rencontrer en français L2, pour le Corpus 2, nous souhaitons nous focaliser sur certains phénomènes en particulier et contrôler finement le contexte dans lequel ils peuvent apparaître. Cette démarche suppose alors que nous « choisissons » certains phénomènes plutôt que d'autres puisque suite à l'enregistrement du Corpus 1, nous avons remarqué qu'il

est préférable que la durée d'enregistrement totale ne dépasse pas une heure par locuteur (ceci pour éviter les effets de fatigue vocale, particulièrement pour les débutants). C'est pourquoi nous avons ajusté la taille globale de notre Corpus 2 tout en conservant le découpage en quatre tâches. Dans les paragraphes suivants, nous allons présenter ce Corpus selon ce découpage.

(1) Tâches de lecture et de répétition (Tâches 1 et 2)

Dans la mesure où les tâches de lecture et de répétition ne posaient pas de problème particulier à nos locuteurs, le nombre de phrases pour ces tâches a été légèrement augmenté pour pouvoir investiguer plus en détails les phénomènes phonétiques et phonologiques sur lesquels nous souhaitons concentrer nos futures études. Le Corpus 2 a été pensé de façon à inclure les phénomènes phonétiques déjà ciblés dans le Corpus 1.

Certains phénomènes ont fait l'objet d'un protocole précis, c'est notamment le cas pour la production des occlusives pour lesquelles nous avons proposé des paires quasi-minimales dans la mesure où le français le permettait (un ajustement a été fait si nécessaire en essayant de conserver un environnement phonétique et syntaxique le plus proche possible). Pour ces phrases, l'environnement vocalique, la position dans la phrase ainsi que la structure syntaxique ont été hautement contrôlés. Deux contextes ont été retenus : occlusive + [oR] et occlusive + [aR]. L'occlusive est l'une des six occlusives du français [b d g p t k]. Rappelons qu'en français, le contraste phonologique est phonétiquement réalisé par le voisement alors qu'en allemand celui-ci est principalement réalisé par l'aspiration, et donc par un bruit d'explosion (Voice Onset Time ou VOT) plus long (voir par exemple : Lisker et Abramson, 1964 ou Cho et Ladefoged 1999).

Le bateau à vapeur a viré de **bord** dimanche.
Le bateau à vapeur a quitté le **port** de Nice.
Le silence se fait quand elle **dort** l'après-midi.
Les accusés ont toujours **tort** dans ce cas.
La police a découvert le **gosse** dans le parc.
La police a découvert le **corps** à terre.

Le parrain a quitté le **bar** ce matin.
Le garçon mange deux **parts** de gâteau.
L'abeille n'a pas de **dard** sur le corps.
Les poids ont fait la **tare** sur la balance.
Le train a quitté la **gare** de Paris.
Le garçon a pris le **car** à Berlin.

L'analyse du Corpus 1 ayant mis en évidence différentes stratégies d'autocorrection en ce qui concerne la production des obstruantes en position finale selon que le locuteur est en tâche de lecture ou de répétition, nous avons choisi de contrôler finement ce phénomène dans le Corpus 2. Aussi, avons-nous retenu le contexte [a] + obstruante, où l'obstruante est l'une des six obstruantes voisées du français [b d g v z □], soit 6 phrases pour la tâche de lecture et 6 pour la tâche de répétition. Le contexte syntaxique et le vocabulaire ont été, pour ce phénomène également, finement contrôlés. Nous avons notamment veillé à ce que l'obstruante soit en position finale dans le mot, mais pas dans la phrase afin qu'il n'y ait pas de relâchement articuloire souvent observé en fin de phrase et à ce qu'elle soit suivie d'une voyelle non élidable afin d'éviter les phénomènes d'assimilation. Nous supposons également que si l'apprenant, notamment le débutant, ne parvient pas à voiser l'obstruante en position finale, il ne réalise pas pour autant pas sa correspondante non voisée. Il est probable qu'il s'agisse plutôt de dévoisement, c'est pourquoi nous avons également ajouté 3 obstruantes non voisées [f s □] dans le même contexte vocalique et syntaxique, uniquement en tâche de lecture, afin de pouvoir étudier ce type de production également. En effet, rappelons que les obstruantes en position finale dans la syllabe sont sujettes au dévoisement en allemand, mais pas en français (par exemple Wiese, 1996 ou Möbius, 2004). Rappelons que le trait phonologique dit de « voisement » qui oppose les obstruantes voisées aux obstruantes non voisées ne repose pas que sur la caractéristique articuloire de voisement (vibration des plis vocaux). Ainsi une absence de voisement n'entraîne pas systématiquement une confusion entre les deux types de consonnes.

Les idiots chassent les **crabes** en Australie.
Vous aimez la **salade** aux anchois.
Son ami comprend la **blague** en anglais.
Les enfants sont **braves** à l'école.
Les élèves doivent cocher la bonne **case** avec un feutre.
Mon ami a perdu ses **bagages** à la gare.
Marc a pris une **baffe** à cause de Marie.
Maman a acheté une table **basse** orange.
La piscine est couverte d'une **bâche** en plastique.

Le panier est rempli de **crabes** abimés.
Les enfants sont partis en **balade** en forêt.
Maman a perdu sa **baguette** en argent.
Le champagne est rangé dans la **cave** en terre.
Les avions sont rentrés à la **base** après le vol.
Elle habite dans un beau **village** en France.

La prosodie sera principalement étudiée à partir de la lecture du texte des *trois petits cochons*. Dans le Corpus 2, nous avons toutefois également inclus des phrases interrogatives pour lesquelles nous avons pris soin de n'avoir que des sons voisés (sauf bien sûr pour la question en « en ce que » qui ne le permet pas) pour permettre une détection et un suivi plus aisé de la fréquence fondamentale. La plupart de nos phrases sont relativement courtes et simples. Trois phrases plus longues ont été retenues afin d'observer où les apprenants feront leurs pauses et si elles sont sémantiquement acceptables.

Aimez-vous mon ami ?
Est-ce que vous aimez mon ami ?
Vous aimez mon ami ?
Où allons-nous en mai ?
La jeune fille a mis une jolie robe rose faite dans un tissu doux.
La méchante sorcière est le personnage principal du spectacle de ce soir.
Nous devrions porter quelques fleurs à la servante au grand cœur dont vous étiez jalouse.

L'allemand, comme l'anglais, possède des voyelles longues et brèves. Les voyelles brèves sont en général réalisées quand elles précèdent une séquence de deux consonnes ne débutant pas par un /h/, que ces consonnes soient identiques ou non. C'est un contraste phonologique, Pollen/Polen (polonais/pollen), Bett/Beet (lit/partir), ou encore offen/Ofen (ouvert/four) par exemple, sont des paires minimales qui ne diffèrent que par la longueur des voyelles (associée à une légère différence de timbre). Les allemands tendent à maintenir ce contraste en français, où il n'existe pas, ce qui contribue fortement à l'impression d'accent étranger. D'autre part, le timbre des voyelles varie avec leur durée. Ainsi, par exemple, la voyelle [o] en allemand est ouverte quand elle est brève -donc quand elle précède deux consonnes-, alors qu'elle est généralement fermée quand elle est longue. Comme on le voit, timbre, quantité et contexte sont étroitement liés en allemand. Afin de tester si le locuteur allemand est influencé par sa langue maternelle en ce qui concerne la réalisation des voyelles, nous avons consacré quelques phrases au timbre vocalique. Dans les exemples suivants, le mot "Rome" est réalisé avec un [o] ouvert en français, mais sera prononcé avec un [o] fermé par un locuteur allemand qui conserve les règles de sa langue maternelle. Ceci pourrait être le cas avec toutes les voyelles semi-ouvertes du français standard. Ce phénomène est notamment investigué dans les phrases suivantes :

J'ai acheté mon armoire à **Rome**.
Ils ont mangé une **pomme**.
La voiture est **belle**.
Les élèves cochent la bonne case avec un **feutre**.

Enfin d'autres phénomènes, comme la réalisation du phonème [R] (que les locuteurs L2 allemand pourraient vocaliser après une voyelle dans la même syllabe) ou de [s] devant [p] (que les locuteurs L2 pourraient prononcer comme [H] en position initiale) par exemple ne sont pas investigués de façon systématique mais restent présents dans le Corpus 2. En parole connectée, il n'est pas rare d'observer des

réductions et des assimilations, particulièrement dans productions moins formelles (voir par exemple Racine 2008), ci-après, quelques exemples de phrases qui sont susceptibles de mettre en lumière ce type de déviations :

Il boit une bière au **bar**.
Marc fait de la poutre en cours de **sport**.
Je ne sais pas si le professeur est absent.
Je n'ai pas le temps.

Nous avons également ajouté d'autres entités linguistiques, non présents dans le Corpus 1, que nous souhaitons tester comme les nombres ou les abréviations. Ceci devrait augmenter la charge cognitive de ces occurrences. Ces phrases devraient alors présenter des variations segmentales intéressantes surtout pour les nombres qui présentent des structures très différentes en allemand et en français. Par exemple, en allemand, le nombre 96 se prononce littéralement « 6 et 90 » alors qu'en français il se prononce « 4 20 16 ». Ces phrases pourraient également être intéressantes sur le plan de la prosodie, en voici quelques exemples :

Cent moins deux est égal à quatre-vingt-dix-huit.
La BMW bleue est immatriculée AM 525 YY.
Barack Obama est le 44ème président des USA.
En 2013, il y a aura 26 millions d'habitants dans l'UE.
Les 5 anneaux sur le drapeau du CIO sont le symbole des 5 continents.

Les mots dits « transparents » et les noms propres devraient nous permettre d'observer les stratégies de l'apprenant et de voir si dans ce contexte phonétique particulier, il utilise la prononciation de la L2 ou de sa L1. Ceci devrait permettre de mettre à jour des prononciations différentes, il est probable que dans ce cas également, les erreurs soient différentes selon le niveau de compétence linguistique du locuteur. Dans les exemples ci-dessous, nous avons souligné les mots transparents et noms propres.

En cette **saison**, Paul commande souvent du **gratin** au **restaurant**.
Mon **cousin** a beaucoup de **chance**, il habite dans un grand **pavillon**.
Le garçon prend le car à **Berlin**.
Le train a quitté la gare de **Paris**.
J'ai acheté mon armoire à **Rome**.

(2) Tâche de focus (Tâche 3)

La tâche de focus a été simplifiée pour le Corpus 2. Le nombre de phrases dans la tâche de focus a été limité à deux (au lieu de six dans le Corpus 1). Ceci permet non seulement de réduire la taille globale du Corpus 2, mais également de ne pas mettre le locuteur dans une situation trop artificielle. A partir de ces deux phrases, le locuteur se voit poser trois questions. Les deux types de focus, contrastif et non contrastif ont été conservés. Pour la constitution de cette tâche, une attention particulière a été portée à la détection et au suivi automatique de la fréquence fondamentale, qui est un indice primordial dans ce type de tâche. Aussi les phrases retenues, sont non seulement des phrases simples en termes de structure syntaxique et de vocabulaire, mais ont également la particularité de ne présenter que des sons voisés et des consonnes sonantes, ce qui devrait faciliter la détection automatique de la fréquence fondamentale en évitant les décrochages. Ceci n'avait pas été contrôlé pour la préparation de cette tâche dans le Corpus 1.

Exemple de corpus non contrastif

Qui met un nœud ?
MARIE met un nœud.

Que met Marie ?
Marie met un NOEUD.

Que fait Marie ?
Marie met un nœud.

Exemple de corpus contrastif :

Marc amène un ami ?
YVONNE amène un ami.

Yvonne amène ma mère ?
Yvonne amène un AMI.

Yvonne dépose un ami ?
Yvonne AMENE un ami.

(3) Tâche de lecture d'un texte (Tâche 4)

En ce qui concerne le texte, nous n'avons retenu qu'une histoire, celle des trois petits cochons parce qu'elle est connue des locuteurs allemands et français, que son style est simple et que les mots sont répétés plusieurs fois ce qui permet de tester si les productions déviantes sont constantes. La réduction à un seul texte permet également de réduire la durée totale d'enregistrement puisque les locuteurs ont souvent eu besoin de lire plusieurs fois les histoires en lecture silencieuse avant leur enregistrement pour le Corpus 1.

2.2 Sujets et enregistrements

Nous souhaitons enregistrer un total de 100 sujets pour le Corpus 2 (voir Table 2), ce qui devrait comprendre des étudiants allemands considérés comme débutants (niveau A2), d'autres considérés comme avancés (niveau C1 ou C2) et enfin des lycéens (niveau A2). Des étudiants et des lycéens francophones natifs seront également enregistrés. Le genre des locuteurs devrait être équilibré dans chacun des groupes.

Table 3: Groupes de sujets qui seront enregistrés pour le Corpus 2 en fonction de leur L1 et de leur L2 (F=Français, A=Allemand), de leur niveau en L2 et de leur âge.

# sujets	L1	âge	# sujets	L1	L2	niveau	âge
40	F	18-30 ans	20	A	F	débutants	18-30 ans
			20	A	F	avancés	
10	F	15-16 ans	10	A	F	débutants	15-16 ans

Les conditions d'enregistrement du Corpus 1 se sont révélées satisfaisantes ; les moyens techniques employés pour le Corpus 2 seront donc les mêmes que celles utilisées pour l'enregistrement du Corpus 1. Toutefois, afin de mieux contrôler encore les effets de fatigue vocale et les effets de liste, les phrases seront proposées dans un ordre aléatoire.

2.3 Annotation et alignement automatique

La totalité du Corpus 2 sera automatiquement segmentée et alignée par notre système d'alignement automatique. La correction manuelle de l'alignement du Corpus 1 va permettre d'améliorer notre outil d'alignement automatique à deux niveaux. En ce qui concerne les prononciations, l'analyse par l'outil CoALT (Fohr & Mella, 2012) des différences entre les alignements automatiques et corrigé du Corpus 1 va extraire des variantes de prononciation fréquemment réalisées par les locuteurs non natifs. Il sera donc possible d'ajouter au lexique les variantes les plus fréquentes, soit en les introduisant directement dans le lexique, soit en dérivant des règles qui permettent de générer automatiquement ces variantes non natives. En ce qui concerne les modèles acoustiques, le Corpus 1 permettra d'adapter ces modèles afin d'être plus proches des réalisations non natives. Cette adaptation est réalisable au moyen des algorithmes MLLR (Maximum Likelihood Linear Regression) ou MAP (Maximum A Posteriori).

3 Conclusions

Un corpus de parole en français prononcé par des locuteurs natifs et par des apprenants L2 est un bon point de départ pour conduire des analyses phonétiques et phonologiques détaillées sur les aspects segmentaux mais également sur les aspects prosodiques de l'apprentissage du français L2 par des locuteurs allemands. Notre corpus se veut finement contrôlé ce qui devrait nous permettre d'isoler les difficultés que les apprenants allemands peuvent rencontrer en français langue étrangère et ainsi pouvoir proposer à terme des corrections individualisées.

Grâce à notre cohorte de locuteurs qui comprend deux niveaux d'apprentissage (A2 et C1) mais également deux classes d'âge (adolescents et adultes), il devrait également être possible d'analyser les différentes erreurs que les apprenants produisent en fonction de leur niveau de compétence en L2 et d'étudier les éventuelles stratégies d'auto-correction qu'ils peuvent mettre en place. Ceci devrait notamment être rendu possible grâce aux deux modalités selon lesquelles l'apprenant lit ou répète une phrase.

De plus, un tel corpus, enrichi en annotations phonétiques fines est une ressource importante comprenant des informations pertinentes pour le public intéressé comme par exemple les enseignants en langues étrangères. Ce corpus est aussi la condition préalable pour le développement d'un logiciel d'apprentissage des langues assisté par ordinateur en adaptant le contenu des commentaires et des exercices pour les apprenants en langue étrangère.

Ce corpus de parole en français prononcé par des locuteurs natifs et des apprenants L2 enrichi d'un alignement automatique manuellement corrigé pourra être mis à la disposition de la communauté scientifique à la fin du projet.

Remerciements

Ce projet a été financé par une ANR/DFG "IFCASL" attribuée à l'équipe Parole du LORIA CNRS UMR 7503 – Nancy France and à l'équipe de Linguistique Computationnelle FR4.7, Université de la Sarre – Saarbrücken Allemagne, 2013 – 2016.

Références bibliographiques

- Adell, J., Bonafonte, A., Gómez, J. A., & Castro, M. J. 2005. Comparative study of automatic phone segmentation methods for TTS. In *Proceedings of ICASSP*, pp. 309-312.
- Andreeva, B. and Barry W. (in press). Fine phonetic detail in prosody. Cross-language differences need not inhibit communication. In: O. Niebuhr and H. Pfitzinger (eds.), *Prosodies, context, function, and communication*. Berlin: de Gruyter.
- Bartkova, K. and Jouvét, D. (2007). On using units trained on foreign data for improved multiple accent speech recognition. *Speech Communication* 49, pp. 836–846.
- Barry W.J., Andreeva, B. and Steiner I. 2007. The Phonetic Exponency of Phrasal Accentuation in French.
- Bonneau, A. and Colotte, V. (2011). Automatic feedback for L2 prosody learning. In Intech "Speech Technology". Available from: <http://www.intechopen.com/books/speech-and-language-technologies/automaticfeedback-for-l2-prosody-learning>.
- Bonneau, A. and Laprie, Y. (2008). Selective acoustic cues for French voiceless stop consonants. *Journal of the Acoustical Society of America* 123, pp. 4482-4497.
- Best, C. T. 1995. A direct realist view of cross-language speech perception. *Cross-language studies of speech*.
- Bouselmi, G., Fohr, D. & Illina, I. (2011). Multilingual recognition of non-native speech using acoustic model transformation and pronunciation modeling. *International Journal of Speech Technology* 15(2), pp. 203-213.
- Cho, T., and Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages, *J.Phonet.* 27, 207–229.

- Cucchiari, C., Strik, H. & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America* 107 (2), pp.989-999.
- Derwing, T.M. and Munro, M.J. (2001). What speaking rates do non-native listeners prefer? *Applied Linguistics* 22, pp. 324-337.
- van Doremalen, J., Strik, H. & Cucchiari, C. (2009). Optimizing non-native speech recognition for CALL applications. Proc. Interspeech, Brighton, pp. 592-595.
- Dupoux, E., Peperkamp, S. and Sebastián-Galles, N. (2001) A robust method to study stress "deafness". *Journal of the Acoustical Society of America*. 110, pp. 1606-1618.
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E. & Peperkamp, S. 2008. Persistent stress 'deafness': The case of French learners of Spanish. *Cognition* 106.682-706.
- Fauth C., Bonneau A. (2014). L1-L2 interference: the case of devoicing of French voiced obstruents in final position by German learners - Pilot study. International Workshop on Multilinguality in Speech Research: Data, Methods and Models. Dagstuhl - Germany
- Flege, J. E. 1988. Effects of speaking rate on tongue position and velocity of movement in vowel production. *Journal of the Acoustical Society of America* 84.901 – 916.
- Flege, J.E. (1995). Second Language Speech Learning: Theory, Findings and Problems. In: Strange, W. (ed). *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, pp. 233-272. Timonium: York Press.
- Flege, J. E., MacKay, I. R. A. & Meador, D. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106.2973-2987.
- Fohr, D. and Mella, O. 2012. CoALT: A software for comparing automatic labelling tools, *Proceedings of LREC*.
- Galliano S., Geoffrois E., Gravier G., Bonastre J-F., Mostefa D., Choukri K. (2006) Corpus description of the ESTER Evaluation Campaign for the Rich Transcription of French Broadcast News. *Proc. Language Evaluation and Resources Conference, 2006*.
- Goronzy, S., Sahakyan, M. & Wokurek, W. (2001). Is non-native pronunciation modeling necessary? Proc. Eurospeech, Aalborg, pp. 309-312.
- Gravier G., Adda G., Paulsson N., Carre M., Giraudel A., Galibert O. (2012). The ETAPE corpus for the evaluation of speech-based TV content processing in the French language. *Processing Language Evaluation and Resources Conference*.
- Gut, U. (2009). Non-native Speech. A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German. (*English Corpus Linguistics* 9). Frankfurt: Peter Lang.
- Hirschfeld, U. and Trouvain, J. (2007). Teaching prosody in German as a foreign language. In: Trouvain, J. & Gut, U. (eds) *Non-Native Prosody. Phonetic Description and Teaching Practice*. (Trends in Linguistics. Studies and Monographs [TiLSM] 186) Berlin/New York: Mouton de Gruyter. 171-187.
- Jilka, M. 2007. Different manifestations and perceptions of foreign accent in intonation. *Non-Native Prosody- Phonetic Description and Teaching Practice*, ed. by J. Trouvain & U. Gut, 77-96. Berlin: Mouton De Gruyter.
- Jilka, M. (2009). Talent and proficiency in language. In G. Dogil and S. Reiterer (eds.). *Language Talent and Brain Activity*. Mouton De Gruyter, Berlin, pp. 1 – 16.
- Jouvet, D., Mesbahi, L., Bonneau, A., Fohr, D., Illina, I. & Laprie, Y. (2011). Impact of pronunciation variant frequency on automatic non-native speech segmentation. Proc. 5th Language & Technology Conference (LTC'11), Poznan, pp. 145-148.
- Kingston, J. 2003. Learning foreign vowels. *Language and Speech* 46.295-349.
- Lisker L. et Abramson A.S. (1964) A cross-language study of initial stops. *Acoustic measurements*. *Word* 20, 384-422.
- Mehlhorn, G. and Trouvain, J. (2007) Sensibilisierung von Lernenden für fremdsprachliche Prosodie. *Zeitschrift für interkulturellen Fremdsprachenunterricht* 12 (2). 25 pp.
- Möbius, B. (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica* 38 (1-2), pp. 5-26.

Piske, T., MacKay, I. R. A. and Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, pp. 191-215.

Toledano, D. and Gomez, L. 2003. Automatic Phonetic Segmentation. *IEEE Trans. on Speech and Audio Processing*, v11, n6, pp. 617—625.

Racine, I. (2008). Les effets de l'effacement du Schwa sur la production et la perception de la parole en français. Thèse de doctorat: Univ. Genève.

Rasier, L., Hiligsmann, Ph., Caspers, J. and van Heuven, V. (2011). Accentual marking of information status in Dutch and French as foreign languages: Production and perception. In: Wrembel, M., Kul, M., Dziubalska-Kolaczyk, K. (Ed., 2010), *Achievements and Perspectives in SLA of Speech: New Sounds 2010, Volume 2* (Polish Studies in English Language and Literature 32) , pp. 227-238. Frankfurt am Main: Peter Lang.

Strik, H., Truong, K., de Wet, F. and Cucchiari, K. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication* 51, pp. 845-852.

Wiese, R. (1996) *The Phonology of German*, Oxford: Clarendon Press.

Witt, S. and Young, S., (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech communication* 30(2-3), pp. 95-108.

Witteman, M. J., Weber, A., and McQueen, J. M. (2011). On the relationship between perceived accentedness, acoustic similarity, and processing difficulty in foreign-accented speech. *Proc. Interspeech*, Florence, pp. 2229-2232.

Zimmerer F., Jügler J., Andreeva B., Möbius B and Trouvain J (2014) Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. Proceedings of the 7th Speech Prosody Conference, Dublin.

<http://www.etsglobal.org/Fr/Eng/Research/CEFR>

<http://raweb.inria.fr/exploraweb/static/2011/parole/uid63.html>

<http://www.uclouvain.be/en-cecl-lcworld.html>

<http://www.phon.ucl.ac.uk/home/sampa/>

<http://www.ipds.uni-kiel.de/forschung/kielcorpus.de.html>

<http://www.fon.hum.uva.nl/praat/>