REPÉRAGE AUTOMATIQUE D'UNITÉS ACOUSTICO-PHONÉTIQUES POUR L'ENSEIGNEMENT DE LA PRONONCIATION FRANÇAISE**J.F. MALET, G. PÉRENNOU* et N. VIGOUROUX****Department of Foreign Languages, California State Univ. sacramento, USA***Dept. Interface Homme-Machine, Parole et Texte, IRIT-URA 1399, CNRS, 118 route de Narbonne, F-31062 Toulouse cedex, France*

Abstract: Basically, we are proposing an automatic segmentation of the vocal signal, in order to tackle the learner's output assessment problem (As far as non-native pronunciation is concerned right now, we are working on samples from learners of French as a Foreign Language). The approach is underpinned by observations, collected within the instructional realm, and which point the advantages of assessing pronunciation at several levels (phonetic features, phonemes). In this paper, we are analyzing the outputs of the segmenting process on the (learner-produced) items as well as on corresponding (teacher-produced) ones and compare them.

Dans une perspective de développement d'un système d'apprentissage, d'évaluation et de corrections de la prononciation en langue étrangère, il est nécessaire d'une part, de faire l'analyse des traits caractéristiques de la parole des locuteurs ne parlant pas leur langue maternelle et d'autre part, de proposer des méthodes de correction (Léon, 1964), (MacKay, 1978) et par suite de développer des outils d'aide à la correction de la prononciation. Dans cet article, nous présentons une méthode d'analyse automatique de la prononciation française par des anglophones d'origines variées pour en déduire un sous-ensemble de réalisations phonétiques, spontanément perçues d'ordinaire comme autant de marques ou de traces d'accents non-francophones. Au-delà du difficile problème de caractérisation (dû au socio-culturel, au psychologique, aux consignes pédagogiques, etc) de ces traits de parole —sélectionnés par l'expert pédagogique pour leur importance dans la fonction évaluative— nous nous intéressons à la localisation exacte, dans la chaîne parlée, des phénomènes acoustico-phonétiques qui en sont porteurs, et nous en recherchons une description aussi invariante que possible.

1. Méthode de détection des unités acoustico-phonétiques

Les manifestations de parole précises, que nous considérons, peuvent se produire soit au niveau du phonème —car plusieurs phonèmes de l'anglais n'ont guère d'équivalent, voire aucun, en français— soit au niveau de segments infra-phonémiques responsables de traits typés (Voir Paragraphe 2. Les unités mises en cause). Elle peuvent en outre se trouver régulièrement associées à des cadres prosodiques connus. Pour nous, tous ces phénomènes, quel que soit le niveau auquel ils sont produits, sont interprétables sur l'axe syntagmatique en termes d'événements acoustico-phonétiques générés automatiquement par le système

SAPHO (Kabré, 1991). Le système SAPHO procède en deux phases : une phase de segmentation et une phase d'identification. La segmentation est effectuée sur une représentation temporelle du signal vocal, à savoir l'amplitude et le taux de passage à zéro déterminés sur une fenêtre d'analyse de quatre millisecondes. De nombreuses expériences (Pérennou, 1991) ont montré la bonne adéquation de ces frontières automatiques —déterminées par rapport aux variations de ces deux seuls paramètres— par rapport à celles d'un étiquetage manuel (90% à 20 ms). Elles ont également montré que ce système est indépendant des corpus, des locuteurs et des langues (ceci a pu être montré en utilisant des corpus de parole continue, multilocuteurs anglais, italiens, français et suédois). Dans un deuxième temps, ces segments sont identifiés en termes de classes phonétiques majeures —eg, vocalique, liquide aigu, liquide grave, etc. On pourra se reporter à (Pérennou, 1991) pour la signification des étiquettes.

Ces expériences attestent de la robustesse et de la qualité de notre méthode de segmentation et d'étiquetage en événements phonétiques. Ceci nous autorise à les utiliser comme éléments de base dans la recherche de traits liés à la prononciation d'une langue étrangère par des apprenants. Pour réaliser cette étude nous disposons d'une base de données de quarante phrases enregistrées par cinq américains apprenant le français et de deux locuteurs français de référence.

2. Les unités phonétiques mises en cause

Nous avons recensé ici quelques unités phonétiques modifiées par les apprenants de la prononciation en langue étrangère. Ces modifications sont identifiables par notre système automatique d'étiquetage en événements phonétiques.

2.1. Voyelles (durée anormale de réalisation)

La durée de réalisation d'une voyelle peut faire l'objet d'une surveillance suivie et d'une comparaison par rapport à un modèle; ce qui fait qu'une consigne de durée maximale acceptable peut être imposée afin de rétablir un tempo d'élocution proportionné. Par exemple, dans la séquence "il a dit", la réalisation du second [i] par l'apprenant rappelle le [i:] anglais. Elle représente les 3/4 de la durée syllabique (Fig.1) contre la 1/2 d'une réalisation francophone proportionnée.

Ce contrôle de durée est également particulièrement utile dans les séquences telles que voyelle+[R]+occlusive, où les sujets anglophones présentent une forte tendance à escamoter le [R] en allongeant démesurément la voyelle. Cette modification se traduit aussi au plan fréquentiel : nous n'en discuterons pas ici car ceci ne peut être considéré dans notre système.

2.2. Voyelles normalement pures, réalisées avec un double noyau vocalique

Egalement détectable, la tendance à la diphongue, alors que le français parlé en France n'est pas censé en recéler. Ce genre de réalisation se traduit généralement au niveau de la segmentation-étiquetage automatique par la présence de deux noyaux vocaliques au lieu d'un seul. A titre d'exemple dans la séquence "vie" réalisée [vi:] on peut voir d'abord un segment-voyelle long et complexe étiqueté WW suivi d'un autre segment-voyelle long étiqueté KK. A noter qu'en anglais américain, les mots empruntés au français subissent fortement cette déformation (eg, "déjà vu" régulièrement réalisé [deiZAvU]).

2.3. Phonèmes et segments occlusifs

La détection des segments occlusifs (étiquettes Q, QQ) par notre système de segmentation en événements ne pose aucun problème. Ce qui fait qu'au niveau phonémique les réalisations, par des

apprenants anglophones, de [tS] pour [S] et de [dZ] pour [Z] sont très facilement détectées, à cause de la présence non ambiguë de segments occlusifs respectivement sourd ou voisé, suivant le cas. Mais cette détection peut s'étendre, au-delà, à toute réalisation non souhaitée de phonèmes occlusifs; donc, de manière typique, en fin de mot ou de syllabe: eg, au "t" et au "d", mais aussi au "p" et au "b", ainsi qu'au "g" et au "c" et plus généralement à toutes lettres orthographiées mais ne devant pas être prononcées en fin de mot. A titre d'illustration, on peut voir Fig. 2, la prononciation de "C'est la ..." réalisée [sEtla] par l'apprenant.

Enfin, notre système devrait permettre du même coup de résoudre le problème de la baisse subite d'énergie sur le [R] ou sur les consonnes après [R] en fin de syllabe (eg, "jour" réalisé [Zu] au lieu de [ZuR] ou encore "porte-manteaux" [pOrma-to] (ou pire [pU@rment@U] au lieu de [pORtma-to]).

Les apprenants anglophones rencontrent, pour la plupart, des problèmes au niveau de l'assimilation phonétique de certaines consonnes; particulièrement lorsque [s] ou [z] sont mis en jeu. Lorsqu'il est orthographié, le "s" est pratiquement toujours réalisé par les débutants et, pour eux, les problèmes abondent en français au niveau du pluriel et de la conjugaison.

D'autres problèmes ont lieu au niveau de la neutralisation entre les traits +voisé et -voisé (eg, "absorbe" réalisé [abzORb] au lieu de [apsORb]). Comme notre système est capable de distinguer entre les sourds et les voisés, ces problèmes d'assimilation peuvent être pris en compte.

2.5. La détection du [@]

Un aspect tactique majeur de l'enseignement de la prononciation française concerne le [@]. Ce phonème ne pose, en soi, pas de problème de réalisation à des anglophones, mais c'est surtout l'application des règles d'interdiction de la réalisation qui créent d'importantes difficultés. Il pourrait paraître a priori moins risqué d'exiger des apprenants la réalisation systématique du [@]. En fait, c'est la tactique inverse qui s'impose; car, outre qu'on ne peut entreprendre d'imposer un trait régional de parole isolé, les réalisations produites par des apprenants américains anglophones, même s'ils ne sont pas hispanophones, débouchent le plus souvent sur [e] plutôt que sur [@].

Fort heureusement, la détection du [@] par notre système est extrêmement aisée, car la réalisation de ce phonème entraîne toujours la présence d'un segment de la classe phonétique voyelle. Il faut noter que [@] n'est pas le seul phonème à être réalisé par [e]; car [œ] et [ø], qui lui sont proches au plan articulatoire, subissent tout aussi fréquemment la même erreur. Cependant la détection de l'erreur [e] pour [œ] ou pour [ø] est un problème relevant plutôt du domaine fréquentiel.

2.6. L'épenthèse en contexte occlusive+liquide

Les rencontres [t]+[R] à l'attaque, et [p ou b]+[l] en finale, semblent poser d'énormes problèmes de réalisation à bon nombre d'apprenants anglophones américains; sans doute par faute de contrôle articulatoire dans la réalisation du [R] et du [l] français. Toujours est-il qu'il y a une forte tendance à intercaler un segment-voyelle épenthétique entre l'occlusive et la liquide. Cette erreur de réalisation est facilement repérée par notre système de segmentation en événements phonétiques.

L'éventail des segments phonétiques détectables par le système SAPHO (segments vocaliques graves ou aigus, segments occlusifs et fricatifs sourds ou voisés, etc) permet de repérer la réalisation, ou la non-réalisation, de segments phonétiques variés, également intéressants au plan de l'enseignement (eg, dans "là-bas" réalisé [laBva] par une anglophone californienne, un segment-voyelle faible reflète la coloration hispanophone du [b], par ailleurs nettement occlusif en français).

Fig. 1 Réalisation de "di"

Fig. 2 Réalisation de "C'est la .."

3. Conclusion et ébauche de la maquette d'exercices

- (1) Etiquetage manuel en phonème
(2) Etiquetage automatique en événements
(3) Signal vocal

Les événements phonétiques obtenus à la sortie de SAPHO sont suffisamment pertinents, et la gamme d'unités phonétiques susceptibles d'être traitées par notre méthode de segmentation (complémentaire des méthodes de traitement de la parole dans le domaine fréquentiel (Lefèvre, 1991)) nous semble assez substantielle pour aborder tout programme de recherche sur l'enseignement automatique de la prononciation française. Ceci nous permet d'envisager une série d'exercices d'évaluation et de correction de la prononciation dans lequel l'apprenant anglophone doit être graduellement amené à prononcer correctement selon un tempo acceptable des phrases françaises. La figure suivante donne les différentes composantes d'un système de contrôle automatique d'exercices de prononciation organisé autour du module VERIPHONE (Pérennou, 1989) qui effectue l'alignement des sorties de SAPHO sur une transcription phonétique donnée.

REFERENCES BIBLIOGRAPHIQUES

- [1] MacKay I.R.A., in Little, Brown & Co, Boston, 1978.
- [2] Léon P. é M., in Hachette et Larousse, 1964 (réédité).
- [3] Kabré H., Pérennou G., Vigouroux N., EUROSPEECH 2nd, Vol.2, (1991), pp. 689-692.
- [4] Pérennou G., Kabré H., Vigouroux N., SAM-Report, (1991).
- [5] LEFEVRE J.P. dans Génie Linguistique, Vol.1 (1991).
- [6] Pérennou G., de Calmès M., Pécatte J.M., Vigouroux N., ESCA Workshop, Vol.1 (1989).&