

Reconnaissance Automatique de la Parole pour des Langues peu Dotées : Application au Vietnamien et au Khmer

L. Besacier (1), V.-B. Le (1), E. Castelli (2), S. Sethserey (3), L. Protin (3)

(1) Laboratoire CLIPS-IMAG, UMR CNRS 5524, BP 53,
38041 Grenoble Cedex 9, FRANCE
(*Laurent.Besacier, Viet-Bac.Le*)@imag.fr

(2) International Research Center MICA, 1 Dai Co Viet,
Hanoi, VIETNAM
Eric.Castelli@mica.edu.vn

(3) Institut de Technologie du Cambodge, Bd de Pochentong
BP 86 – Phnom Penh, CAMBODGE
Sam.Sethserey@itc.edu.kh , Ludovic.Protin@online.com.kh

Mots-clés : reconnaissance automatique de la parole, langues mal dotées, modèles multilingues, ressources écrites et orales multilingues.

Keywords: automatic speech recognition, under-resourced languages, multilingual models, multilingual text and speech resources.

Résumé Nous présentons dans cet article une méthodologie qui vise à développer et adapter le plus rapidement possible un système de reconnaissance automatique de la parole continue pour une nouvelle langue peu dotée. Les ressources collectées et les résultats expérimentaux obtenus pour le vietnamien sont présentés. Notre meilleur système pour cette langue obtient actuellement un taux de reconnaissance de mots de 64% environ. Les travaux en cours sur la langue khmère sont également décrits à la fin de cet article.

Abstract We present here a methodology for fast development of ASR systems for new under-resourced languages. The resources collected for vietnamese, and the experimental results of our first vietnamese ASR system are presented. Our best system obtains 64% of word accuracy rate. The current validation of our methodology for khmer language is also described at the end of this paper.

1 Introduction

L'informatique est désormais un instrument pour écrire et communiquer. Traitements de textes, courriers électroniques, voire des systèmes plus avancés comme la dictée ou la synthèse vocale sont des outils largement répandus. La possibilité d'offrir ou pas ces services

pour une langue donnée peut permettre de définir son « niveau d'informatisation » (Berment 2004).

Avant de soumettre un article à un atelier intitulé « TAL et langues peu dotées », il convient d'abord de vérifier que notre travail correspond bien au thème dudit atelier, à savoir : a) est-ce du TAL ? b) les langues traitées sont-elles peu dotées ? Pour cela, la réponse peut venir de (Berment 2004) qui définit dans sa thèse la notion de langue peu dotée, en introduisant un indice σ mesurant la satisfaction des utilisateurs de logiciels et, incidemment, le niveau d'informatisation de la langue pour une classe de services donnée (si cet indice est inférieur à 10, la langue peut être considérée comme peu dotée informatiquement) :

- a) première réponse à notre interrogation, parmi les services mentionnés par (Berment, 2004) intervient le *traitement de l'oral* qui regroupe les technologies de synthèse et reconnaissance vocale. Il est peut-être bon de préciser ici, même si cela est évident, que ces technologies, bien que faisant intervenir des problèmes liés à l'acoustique et au traitement du signal, nécessitent également de résoudre des problèmes de TAL (conversion graphème-phonème pour la synthèse, et modélisation du langage pour la reconnaissance automatique de la parole, par exemple).
- b) seconde réponse à notre interrogation, les langues que nous abordons font effectivement partie de la classe des *langues mal dotées* (bien que le vietnamien soit à la limite), tel que définie par (Berment 2004). Le khmer, évalué dans sa thèse obtient un indice σ d'environ 6/20, tandis que notre évaluation du vietnamien, faite par E. Castelli et Nguyen Quoc Cuong, tous deux chercheurs du centre MICA à Hanoï, donne un indice σ de 10/20 environ. Une partie de ces différences s'explique notamment par le fait que le vietnamien, contrairement au khmer, utilise une écriture latine accentuée, rendant plus « faciles » les tâches de reconnaissance automatique de caractères et de tri par exemple.

Par ailleurs, il est important de noter que pour ces deux langues, les services liés au traitement de l'oral sont inexistantes. C'est ici que se situe notre problématique. Face à la critique qui pourrait être faite concernant notre choix d'aborder des technologies peut-être moins importantes, en terme de développement, que d'autres liées au traitement de texte et aux dictionnaires, nous argumenterons que développer des systèmes de traitement de la parole dans une langue « peu dotée » peut permettre de collecter des ressources utiles pouvant être ensuite remises au « pot commun » d'une langue donnée (dictionnaire phonétique, corpus oral, transcriptions de conversations spontanées par exemple).

Le présent article détaillera d'abord dans la *section 2* le contexte de nos travaux, qui a conduit au choix des deux langues abordées dans ce travail. Notre méthodologie permettant de développer et d'adapter le plus rapidement possible un système de reconnaissance automatique dans une nouvelle langue, sera décrite dans la *section 3*. La *section 4* montrera une première application de cette méthodologie au vietnamien et les quelques résultats associés (ressources obtenues et expériences de reconnaissance vocale). La *section 5* décrit quand à elle nos travaux en cours sur le khmer. Nous précisons toutefois dès cette introduction que nous ne disposons pas encore d'un système de reconnaissance automatique de la parole pour la langue khmère. Nous décrirons cependant dans cette partie les ressources déjà collectées pour cette langue, ainsi que les travaux prévus dans un futur proche. Finalement, la *section 6* conclura ce travail.

2 Contexte

Le Centre de recherche international MICA a été créé en 2002 pour participer au développement des technologies de l'information au Vietnam et pour répondre aux préoccupations relatives à leur évolution. Les travaux de recherche menés à Mica visent à étudier et à développer des résultats théoriques et des applications dans les thèmes du traitement des signaux complexes (parole et image), des applications multimédia et de l'instrumentation. Une collaboration existe déjà depuis la création de MICA en 2002 avec le laboratoire CLIPS sur le traitement de la langue vietnamienne.

Plus récemment, dans le cadre d'un projet financé par l'AUF¹, le Département de Génie Informatique et Communication de l'Institut de Technologie du Cambodge (ITC) s'est associé à cette collaboration. Le but de ce projet est notamment d'initier au Cambodge, la création d'un nouveau groupe de recherche spécialisé en traitement de la parole en langue khmère. Pour cela, le Centre MICA, qui a démarré ses activités en 2002 et possède désormais un groupe de recherche en parole pour la langue vietnamienne, s'est proposé d'aider l'ITC à finaliser cet objectif.

2.1 La langue vietnamienne

Elle est parlée par environ 70 millions de personnes dans le monde (source : MSN-Encarta). Son origine est toujours sujette à débat parmi les linguistes. Il est cependant généralement admis qu'elle a des racines communes et fortes avec le môn-khmer qui fait partie de la branche austro asiatique et qui comprend le mon parlé en Birmanie et le khmer, la langue cambodgienne, aussi bien que les khmu, bahnar et bru, d'autres langues parlées par les habitants des îles du nord du Vietnam. C'est une langue tonale qui possède six tons. L'orthographe est latine depuis le XVII^e siècle, avec des caractères accentués pour les tons.

2.2 La langue khmère

Elle est parlée par une dizaine de millions de personnes dans le monde (source : MSN-Encarta). Elle appartient également au groupe des langues môn-khmères. La langue khmère est une langue atonale – contrairement aux langues chinoises, thaïes ou vietnamiennes. Cependant, le khmer possède comme ses cousines austro-asiatiques plusieurs registres vocaliques : les voyelles peuvent être allongées (dites voyelles longues), raccourcies (dites voyelles brèves), diphtonguées, reposer sur des consonnes aspirées ou non aspirées, ce qui en modifie complètement le sens. (Ex. slap = mourir; slaaap = aile d'un oiseau). Cette particularité fait du cambodgien un des plus riches systèmes vocaliques au monde. Au niveau de l'écriture, pour adapter les fontes informatiques, il a fallu gérer un ordonnancement, sur plusieurs niveaux, de 33 consonnes, 32 consonnes souscrites, 28 voyelles, 14 voyelles indépendantes et 10 ligatures, sans compter les chiffres et la ponctuation.

¹ Projet AUF/TALK : Traitement Automatique de la Langue Khmère

3 Méthodologie

3.1 Rappel des principes généraux de la reconnaissance automatique de la parole

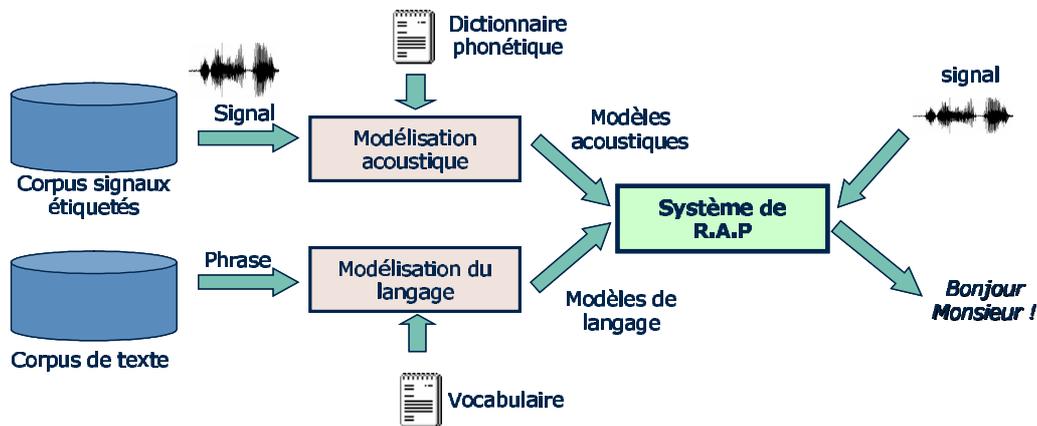


Figure 1. Schéma d'ensemble de la reconnaissance automatique de la parole

Comme l'illustre la *figure 1* ci-dessus, pour développer un système de reconnaissance automatique de la parole continue dans une nouvelle langue, il est souvent nécessaire de rassembler une grande quantité de corpus, contenant à la fois des signaux de parole (pour l'apprentissage des modèles acoustiques du système) mais également des données textuelles (pour l'apprentissage des modèles de langage du système). De tels corpus et systèmes sont désormais disponibles pour la plupart des langues occidentales (anglais, français, espagnol, etc) et pour quelques langues asiatiques (chinois, japonais, etc). Porter un système de reconnaissance vers une nouvelle langue est donc une tâche très fastidieuse si aucun corpus de grande envergure n'existe dans la langue cible, puisqu'il faut alors collecter soi-même les ressources nécessaires : signal de parole, lexicale, corpus textuels, etc. Précisons aussi qu'étant donnée la nature statistique des modèles généralement utilisés en reconnaissance automatique de la parole (modèles acoustiques de phonèmes correspondant à des chaînes de Markov où chaque état est une distribution multigaussienne ; et modèles de langage N-grammes), ces ressources doivent être disponibles en quantité importante.

3.2 Collecte de ressources

Une première façon d'accélérer la portabilité des systèmes de reconnaissance automatique de parole continue grand vocabulaire vers une nouvelle langue, est de développer une méthodologie permettant une collecte rapide et/ou facilitée de ressources textuelles et acoustiques. Cette approche a l'avantage de ne pas modifier fondamentalement le cœur des techniques de reconnaissance utilisées.

Recueil de données textuelles

Concernant le recueil de données textuelles en grande quantité (*figure 2*), une approche intéressante consiste à « aspirer » un grand nombre de sites Web dans la langue donnée et à

filtrer les données récupérées pour les rendre exploitables. Ces données textuelles peuvent servir d'une part à calculer des modèles de langages statistiques, et d'autre part à obtenir un corpus pouvant ensuite être prononcé par des locuteurs en vue de la constitution d'une base de signaux conséquente.

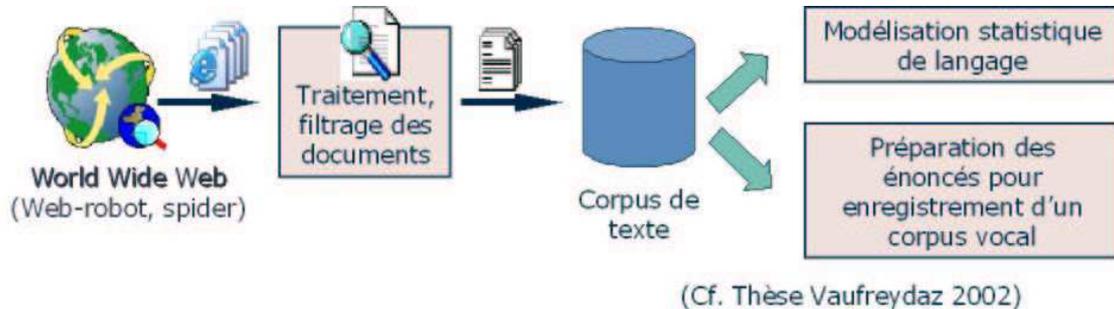


Figure 2 : récupération de données textuelles en utilisant le Web

Une telle approche a déjà été relativement bien validée pour une langue bien dotée telle que le français (Vaufreydaz, 2002). Les problèmes spécifiques pour les langues mal dotées concernent le nombre de sites web qui peut être peu important, la vitesse de transmission et la qualité des documents qui nécessitera alors plus d'outils de traitement. On préférera par exemple des sites de nouvelles, au fort contenu rédactionnel tels que VNexpress² par exemple, pour le vietnamien. Afin de rendre les données exploitables, un certain nombre de traitements sont nécessaires tels que : 1) transformation html vers texte, 2) normalisation des tags, 3) conversion des encodages (nous avons choisi de tout convertir vers une représentation interne unique utilisant l'encodage UTF-8 d'Unicode), 4) séparation en phrases et 5) en mots, 6) groupement de mots composés, 7) transcription des symboles et 8) filtrage en fonction d'un vocabulaire donné. Alors que certains traitements peuvent être considérés comme relativement indépendants de la langue cible (1-2-6-8), d'autres doivent être repensés (3-4-5-7) pour chaque nouvelle langue cible : par exemple la séparation en mots est triviale pour les écritures latines mais problématique pour d'autres systèmes d'écriture comme le khmer, surtout si l'on ne dispose pas d'un vocabulaire (i.e. une liste de mots) au départ. Une boîte à outils *open source* rassemblant quelques uns de ces outils de traitement a été développée au CLIPS³. Une description plus détaillée des traitements réalisés et des expérimentations associées pour la modélisation du langage en vietnamien peut être trouvée dans (Le, 2003).

Recueil de signaux de parole

Pour le recueil de signaux de parole, le CLIPS a développé un outil logiciel ne mettant en œuvre que du matériel standard : EMACOP (Environnement Multimédia pour l'Acquisition et la gestion de Corpus Parole) (Vaufreydaz, 98). La plupart du temps, les campagnes d'enregistrement mobilisent d'importantes ressources humaines pour guider ou assister les locuteurs dans leur tâche de diction, pour organiser l'enregistrement, pour préparer les scénarios et les données, etc. Il faut pouvoir contrôler les différents scénarios pour varier les

² <http://www.vnexpress.net>

³ <http://www-clips.imag.fr/geod/User/viet-bac.le/outils/>

conditions de capture : la lecture d'un texte ou d'une suite de mots ou de mots isolés, la répétition après écoute d'une phrase, le dialogue en réponse à des questions, etc. Les méthodes d'acquisition rigoureusement contrôlées sont donc lourdes et les difficultés sont amplifiées dans le cas des langues mal dotées où les locuteurs ne sont pas forcément rôdés à l'utilisation de moyens informatiques par exemple. C'est pourquoi, le développement d'un utilitaire portable de gestion et d'acquisition de grands corpus sur un matériel standard, nous est d'un grand bénéfice. Le logiciel respecte le format SAM de définition de bases de signaux. Les interfaces ont été adaptées pour manipuler respectivement les caractères vietnamiens et khmers. La figure 3 montre un exemple de l'interface d'EMACOP adaptée pour la langue khmère.



Figure 3 : Interface EMACOP adaptée pour le Khmer

3.3 Dictionnaire de prononciation

Un dictionnaire de prononciation (ou dictionnaire phonétique) est une ressource essentielle aux tâches de synthèse et de reconnaissance de la parole, ou tout simplement pour enrichir un dictionnaire bilingue, permettant au locuteur étranger de connaître la prononciation du mot en langue cible. Cette tâche est cependant difficile pour des langues mal dotées dont le système phonologique est parfois méconnu, ou sujet à débats (langues peu ou mal décrites). Si nous mettons de côté les méthodes manuelles de phonétisation qui, bien que donnant les dictionnaires de prononciation de meilleure qualité, ne nous semblent pas entrer dans le cadre de notre méthodologie, on peut distinguer deux types d'approches automatiques pour constituer un dictionnaire phonétique dans une nouvelle langue :

- Des approches à base de règles, qui nécessitent une bonne connaissance de la langue et de ses règles de phonétisation (qui par ailleurs ne doivent pas contenir trop d'exceptions). Ce type d'approche est assez coûteux en temps (écriture d'un analyseur phonétique), mais donnera des dictionnaires de prononciation de qualité très correcte pouvant ensuite être révisés manuellement relativement rapidement.
- Des approches utilisant un système de reconnaissance phonémique appliqué sur des enregistrements des mots à phonétiser, permettant un premier étiquetage automatique en phonèmes d'une liste de mots qui peut être alors révisé par un opérateur humain.

L'avantage d'une telle approche est bien sûr sa rapidité. Ses inconvénients sont qu'elle nécessite l'emploi d'un système de reconnaissance automatique de phonèmes qui sera généralement celui d'une langue source bien dotée (par exemple un système de reconnaissance des phonèmes du français); par ailleurs, l'autre défaut est que les unités phonémiques décrivant les mots en langue cible seront seulement celles pouvant être reconnues par le décodeur en langue source, d'où la nécessité d'employer si possible des décodeurs phonémiques multilingues pour augmenter au maximum la couverture phonémique dans l'alphabet phonétique international (API). La figure 4 illustre ce problème : la langue source utilisée est le français tandis que la langue cible est le vietnamien. Il est évident que la couverture du vietnamien par le français n'est pas du tout optimale. Une telle méthode reste cependant intéressante, notamment lorsqu'on passe d'une langue source à une langue cible qui possèdent un système phonologique proche.

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal	
Plosive	Ⓟ Ⓠ			Ⓟ Ⓠ		Ⓟ Ⓠ	Ⓟ Ⓠ	Ⓟ Ⓠ	q G		ʔ	
Nasal	Ⓟ	mŋ		Ⓟ		ŋ	Ⓟ	Ⓟ	N			
Trill	B			r					R			
Tap or Flap				ɾ		ɽ						
Fricative	ɸ β	Ⓟ Ⓠ	θ ð	Ⓟ Ⓠ	Ⓟ Ⓠ	ʂ ʐ	ç j	x	χ ʁ	ħ ʕ	h ɦ	
Lateral fricative				ɬ ɮ								
Approximant		u		ɹ		ɻ	j	ɰ				
Lateral approximant				Ⓟ		ɭ	ʎ	ʟ				

○

Phonème FR

□

Phonème VN

Figure 4 : couverture phonémique du français et du vietnamien pour les consonnes

3.4 Modèles acoustiques

Nous avons vu au paragraphe 3.2 qu'il existe des solutions pour collecter rapidement des ressources orales et écrites dans une nouvelle langue. Dans l'idéal, si ces ressources sont en grande quantité, et si un dictionnaire de prononciation est disponible pour la langue cible, l'adaptation du système de reconnaissance peut correspondre alors à un simple réapprentissage des modèles sur ces nouvelles données. Dans la réalité, la quantité de données collectées restent bien souvent inférieure à ce qu'elle est pour les langues bien dotées. La construction d'un système de reconnaissance automatique de la parole nécessite donc également des techniques d'adaptation rapide au niveau des modèles acoustiques comme cela est proposé dans (Choukri 2001) et (Schultz 2001) par exemple.

Une approche possible consiste à obtenir un tableau de correspondances phonémiques (*phone mapping*) entre une ou plusieurs langues sources, et la langue cible. Ensuite, les modèles acoustiques des phonèmes en langue source peuvent être dupliqués pour obtenir des modèles acoustiques en langue cible. L'avantage d'une telle approche est qu'elle ne nécessite pas ou peu de signaux d'apprentissage en langue cible puisque les modèles acoustiques du système de reconnaissance en langue cible sont en fait ceux d'une autre langue. Cependant, on retrouve dans cette approche les mêmes défauts que ceux mentionnés dans le deuxième point du paragraphe 3.3, à savoir le problème de la couverture phonémique (i.e. reconnaître du vietnamien avec des modèles acoustiques appris sur du français !). De tels systèmes peuvent cependant être améliorés en adaptant, par exemple, les modèles acoustiques avec une quantité réduite de signaux en langue cible.

Le problème est aussi d’obtenir le fameux tableau de correspondances phonémiques entre langue cible et langue source. Pour cela, on distingue les méthodes manuelles à base de connaissances (*knowledge-based*), des méthodes automatiques (*data-driven*). Les méthodes manuelles consistent à chercher les couples de phonèmes source/cible les plus proches dans le tableau d’API et nécessitent des connaissances acoustiques et phonétiques des deux langues (source et cible). Une approche automatique consiste plutôt à disposer d’un corpus vocal en quantité limitée en langue source et étiqueté (quelques minutes peuvent suffire), puis à utiliser un décodeur phonémique et calculer la matrice de confusion entre les phonèmes reconnus en langue source et les phonèmes de référence en langue cible. Une description plus détaillée des traitements réalisés et des expérimentations associées pour l’adaptation rapide de modèles acoustiques au vietnamien se trouve dans (Le, 2005).

4 Application au vietnamien

4.1 Ressources collectées

Ressources textuelles

La méthodologie décrite dans le paragraphe 3.2 a été appliquée au vietnamien. La quantité de pages Web collectées était de 2.5Go. Après filtrage, la quantité de données textuelles pouvant servir à l’apprentissage d’un modèle de langage statistique était d’environ 400Mo (5 millions de phrases). A titre de comparaison, une année complète du journal Le Monde en français correspond à 120Mo en moyenne.

Base de signaux de parole

Le corpus de parole vietnamien est toujours en cours d’enregistrement à MICA. A ce jour, il contient 35 locuteurs, 16 femmes et 19 hommes, venant des régions nord, centre et sud du vietnam. Chaque locuteur a enregistré environ 1 heure de parole ce qui fait un total de 35heures. Le corpus contient des séquences de lettres, de nombres et de mots isolés, mais aussi la lecture de phrases complètes et de paragraphes.

Des détails supplémentaires sur les ressources collectées pour le vietnamien se trouvent dans (Le, 2004).

4.2 Quelques expériences de reconnaissance automatique du vietnamien

Dictionnaire phonétique

Au début de ce travail, il n’existait à notre connaissance aucun dictionnaire phonétique sous forme électronique pour le vietnamien. Nous avons extrait tout d’abord un vocabulaire de 6,492 mots isolés à partir d’un dictionnaire franco-vietnamien issu du projet Papillon⁴. Ensuite, un analyseur phonétique (VNPhoneAnalyzer, voir Le, 2004) à base de règles a été

⁴ <http://www.papillon-dictionary.org/>

développé pour obtenir automatiquement un dictionnaire de prononciation vietnamien. Ce dictionnaire phonétique a ensuite été vérifié par des experts de l'Institut Linguistique du Vietnam.

Système de reconnaissance automatique du vietnamien

Notre système de reconnaissance utilise la boîte à outils Janus-III de CMU. Nous avons appliqué la méthodologie décrite au paragraphe 3.4 avec le français comme langue source et le vietnamien comme langue cible. Nous sommes conscients que le choix du français comme langue source n'est pas du tout optimal et nous travaillons actuellement à l'utilisation de modèles source multilingues. Nous avons testé deux techniques d'obtention du tableau de correspondances phonémiques (*knowledge-based* et *data driven*). Les performances du système de reconnaissance automatique de parole continue du vietnamien testé sur un corpus d'une heure de dialogues sont présentées dans la *figure 5* avec respectivement l'utilisation de 0h, 1h et 2h de signal en langue vietnamienne pour adapter les modèles acoustiques empruntés au français.

Génération du tableau de correspondances phonémiques	Quantité de signaux d'adaptation en langue cible		
	0h <i>non-adap.</i>	1h	2h
Manuelle (<i>knowledge-based</i>)	16.13	60.4	63.6
Automatique (<i>data-driven</i>)	18.52	61.6	63.8

Figure 5 : performances (%mots corrects reconnus) de notre système de reconnaissance du vietnamien en fonction de la quantité de signaux d'adaptation utilisée et de la méthode de génération des correspondances phonémiques

Ces résultats montrent le potentiel de l'approche automatique pour la génération du tableau de correspondances phonémiques qui donne des performances équivalentes à celle obtenues avec la méthode manuelle. Nous voyons également qu'avec une quantité réduite de signaux en langue cible (2h), il est possible d'obtenir des performances acceptables (63.8% de mots correctement reconnus), même si la couverture phonémique n'est pas optimale (français vers vietnamien).

5 Travaux en cours sur le khmer

Les travaux actuels sur le khmer concernent essentiellement la collecte de ressources. Des ressources textuelles ont été recueillies avec la méthodologie présentée dans cet article et l'enregistrement d'un corpus en langue khmère est en cours à l'ITC.

Ressources textuelles

A l'aide d'étudiants cambodgiens, nous avons cherché un nombre réduit de pages Web publiées par le gouvernement cambodgien, par des organisations ou des compagnies. Nous avons d'abord remarqué que beaucoup de sites web hébergés au Cambodge sont en fait écrits

en anglais ou en français. Il y a cependant quelques sites écrits en langue khmère. Avec ceux-ci, il y a encore des difficultés de récupération automatique: sites écrits en flash⁵, pages encodées par un système d'encodage spécifique ou privé⁶ que nous n'avons pas réussi à convertir en un autre encodage (Unicode). Nous avons trouvé cependant un site des nouvelles en khmer⁷. La quantité de pages Web collectées à partir de ce site est actuellement de 80Mo (environ 6000 pages), ce qui reste faible par rapport aux 2.5Go de pages web collectées pour le vietnamien et par rapport aux 40Go pour le français (corpus WebFR4) (Vaufreydaz, 2002).

Concernant les modèles de langage, nous avons identifié deux pistes possibles : la première consisterait à faire un système de reconnaissance syllabique où les co-occurrences modélisées seraient des suites de syllabes ; la seconde consisterait à faire un système de reconnaissance de mots où les co-occurrences modélisées seraient des suites de mots. Dans les deux cas, se pose le problème de la segmentation (en syllabes ou en mots) qui est difficile pour les langues comme le thai et le khmer. Il existe des travaux sur ces langues pour résoudre ce problème : utilisation de grammaires de syllabes (Berment, 2004), méthodes basés sur un vocabulaire, méthodes probabilistes (Meknavin, 1997), ...

Base de signaux de parole

En appliquant la boîte à outils de traitement de corpus de texte mentionnée dans la section 3.2, nous avons obtenu dans un premier temps un corpus de phrases à prononcer afin d'effectuer des enregistrements. Pour cela, nous avons utilisé un vocabulaire de 16,000 mots pour filtrer des phrases qui seront énoncés. Ce vocabulaire a été obtenu à partir du dictionnaire khmer Chuon Nat⁸. Le corpus de parole khmer est en cours d'enregistrement à l'ITC. A ce jour, 3 heures de parole ont été enregistrées par 6 locuteurs phnom-penhois.

6 Conclusion

Nous avons décrit dans cet article notre méthodologie permettant de développer et d'adapter le plus rapidement possible un système de reconnaissance automatique de la parole pour une nouvelle langue peu dotée. Des résultats expérimentaux ont été présentés pour le vietnamien et une validation sur la langue khmère est en cours. Dans le futur proche, nous travaillerons essentiellement à une meilleure couverture phonémique par l'emploi de modèles acoustiques multilingues.

⁵ <http://www.everyday.com.kh>

⁶ <http://www.seasite.niu.edu/khmer/>

⁷ www.cambodiacic.org

⁸ <http://www.khmeros.info/>

Remerciements

Ce travail a été réalisé avec l'aide de l'agence universitaire de la francophonie (AUF) dans le cadre du projet TALK.

Références

BERMENT V. (2004) Méthodes pour informatiser des langues et des groupes de langues peu dotées. Doctorat de l'Université J. Fourier – Grenoble I, Mai 2004.

CHOUKRI, R. & al. (2001) Portability of Automatic Speech Recognition Technology to New Languages: Multilinguality Issues and Speech/Text Resources. Panel Session on Automatic Speech Recognition and Understanding (ASRU-2001), Madonna di Campiglio, Italy, December 2001.

LE V.-B., BIGI B., BESACIER L., CASTELLI E. (2003), Using the Web for fast language model construction in minority languages, Eurospeech '03, Geneva, Switzerland, September 2003.

LE V.-B., TRAN D.-D., CASTELLI E., BESACIER L., SERIGNAT J.-F. (2004), Spoken and written language resources for Vietnamese, LREC 2004, Lisbon, Portugal, 26-28 May 2004.

LE V.-B., BESACIER L. (2005), First steps in fast acoustic modeling for a new target language: application to Vietnamese, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), Philadelphia, USA, 19-23 March 2005.

MEKNAVIN, S., CHAROENPORNSAWAT, P. KIJSIRIKUL, B., (1997), Feature-based Thai Word Segmentation, In Proceedings of the Natural Language Processing Pacific Rim Symposium 1997 (NLPRS'97), Phuket, Thailand.

SCHULTZ, T., WAIBEL, A. (2001). Language independent and language adaptive acoustic modeling for speech recognition. *Speech Communication*, vol. 35, no. 1-2, pp. 31–51, 2001.

VAUFREYDAZ D., AKBAR M., CAELEN J., SERIGNAT J.-F., (1998) EMACOP Environnement Multimédia pour l'Acquisition et la gestion de Corpus Parole, JEP'98 (Journées d'Étude sur la Parole), Martigny (Switzerland), pp. 175-178, June 1998.

VAUFREYDAZ D., (2002) Modélisation statistique du langage à partir d'Internet pour la reconnaissance automatique de la parole continue. Doctorat de l'Université J. Fourier – Grenoble I, Janvier 2002.