HAL
open science

Détection de séquences par sélection de l'historique : application à la reconnaissance automatique de la parole

David Langlois, Kamel Smaïli, Jean-Paul Haton

To cite this version:

David Langlois, Kamel Smaïli, Jean-Paul Haton. Détection de séquences par sélection de l'historique : application à la reconnaissance automatique de la parole. XXIVe Journées d'Etudes sur la Parole - JEP'2002, Jun 2002, Nancy, France, pp.301. inria-00107575

HAL Id: inria-00107575

https://inria.hal.science/inria-00107575v1

Submitted on 19 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection de séquences par sélection de l'historique : application à la reconnaissance automatique de la parole

David Langlois, Kamel Smaïli, Jean-Paul Haton

LORIA/INRIA
615 rue du jardin botanique 54602 Villers-lès-Nancy FRANCE
Tél. : ++33 (0)3 83 59 20 00 - Fax : ++33 (0)3 83 27 83 19
Mél : {langlois,smaili,jph}@loria.fr - http ://www.loria.fr/∼langlois

RÉSUMÉ

This paper focuses on statistical language modelling for automatic speech recognition. We present a method which aims at finding linguistic units in corpus. This method, called the Selected History Principle, consists in finding strong distant relationships between words. The new units are phrases made up of basic units of our vocabulary linked by these distant relationships. We adapt the multigram principle to large vocabularies in order to introduce an optimal subset of these sequences into a bigram model. The bigram model using these sequences outperforms the basic bigram model by 21% in terms of Perplexity, and increases the recognition rate of the large vocabulary system Sirocco by 8.7%. The word error rate is decreased by 12.7%.

1. INTRODUCTION

L'unité lexicale la plus communément utilisée dans le cadre de la reconnaissance automatique de la parole (RAP) est la forme graphique telle qu'elle doit être affichée par le système. Or, les unités lexicales du langage ne sont pas nécessairement en adéquation avec ces formes graphiques. Notamment, une unité lexicale peut être composée de la suite de plusieurs formes graphiques. Un exemple révélateur de telles unités est « pomme de terre ». Il est important de prendre en compte ces particularités dans le cadre de la modélisation stochastique du langage, composante importante d'un système de reconnaissance. La communauté a ainsi travaillé sur des modèles de séquences [1, 4, 11] qui prennent en compte la notion de séquence en ajoutant au vocabulaire de formes graphiques d'autres unités formées par la concaténation d'unités de base. On aurait ainsi, par exemple, une unité « au_fur_et_à_mesure_que » ou « petits_pois ». L'élaboration de tels modèles demande de déterminer un jeu de séquences utiles en terme de modélisation statistique du langage, d'une part, et l'introduction de ces nouvelles unités dans le vocabulaire et le modèle de base, d'autre part.

Nous proposons dans cet article d'intégrer un jeu de séquences préalablement déterminé à partir du vocabulaire de base. Ces séquences ont été sélectionnées selon une heuristique fondée sur le Principe de Sélection par l'Historique (SHP) [9]. Ce nouveau principe est fondé sur une mesure de la capacité prédictive d'un modèle de langage en fonction de l'historique. Il permet, entre autres, de déceler de fortes relations intra-lexicales, et donc les unités lexicales reposant sur ces relations. Les séquences déterminées sont nombreuses, au delà des capacités d'un système de reconnaissance. Nous avons donc adapté le principe des multigrammes [1] aux modèles n-grammes à grands vocabulaire afin de déterminer un jeu de séquences à la fois de taille raisonnable et utile en terme de description statistique de langage. Ce principe répond à notre problème qui est ici d'intégrer dans un modèle statistique de langage un sous-ensemble utile des séquences obtenues *a priori* grâce au principe SHP.

Par la suite, nous donnons un aperçu des méthodes de construction de modèles de séquences, et introduisons le principe des multigrammes ainsi que son adaptation aux grands vocabulaires. Puis nous décrivons le principe SHP qui permet de proposer un jeu de séquences de taille restreinte à l'initialisation de l'apprentissage des multigrammes. Nous évaluons alors en terme de perplexité le modèle de séquences obtenu et donnons l'apport en terme de performances pour le système de reconnaissance Sirocco [5].

2. MÉTHODES DE CONSTRUCTION DE MODÈLES DE SÉQUENCES

Dans le cadre de nos modèles statistiques de langage, nous avions déjà pris en compte la notion de séquence en ajoutant au sein de nos modèles des unités correspondant à des locutions de la langue Française, déterminées manuellement, telles que « à_moins_que », « tout_de_même », « aujourd_hui »... Nous avons ainsi ajouté 130 de ces locutions à notre vocabulaire de base de 20K mots. Ces formes ont été sélectionnées grâce à nos connaissances linguistiques. Mais, si on souhaite en trouver d'autres qui ne sont pas recensées, il est important d'automatiser le processus de détection de ces séquences. Il existe des méthodes automatiques permettant de sélectionner un jeu de séquences optimal selon un critère donné, à partir d'un vocabulaire de base de formes graphiques isolées. On peut distinguer deux courants : l'un construit les séquences au fur et à mesure de l'apprentissage du modèle de séquences ; l'autre les construit toutes *a priori*, avant tout apprentissage.

Le premier courant [4] consiste à concaténer deux à deux les éléments du vocabulaire et à intégrer dans celui-ci les meilleures séquences selon un critère tel que l'information mutuelle. Le processus est itéré jusqu'à convergence de la perplexité du modèle de séquences sur un corpus de développement. A chaque itération, les corpus d'apprentissage et de développement sont ré-écrits en tenant compte des nouvelles unités. Ainsi, la première étape permet de construire des séquences de longueur 2, qui peuvent être alors utilisées à la deuxième étape pour construire des séquences de longueur 3 ou 4, et ainsi de suite. Ces méthodes sont efficaces, mais ne conviennent pas à notre cas. En effet, notre jeu de séquences est déterminé en utilisant le principe SHP et donc avant toute intégration dans un modèle de langage. De plus nos séquences ne sont pas construites de manière incrémentale, mais chacune est déterminée en une seule étape, quelle que soit sa longueur. C'est donc d'une méthode d'intégration d'un jeu de séquences défini *a priori* dont nous avons besoin.

Le second courant est fondé sur le principe des multigrammes. Ce principe permet de déterminer de manière itérative la participation des séquences au corpus d'apprentissage et de supprimer les séquences trop rares, qui sont plus un poids qu'un atout dans le modèle de séquences final. Le processus est itéré jusqu'à convergence de la perplexité. Cette méthode est non supervisée et repose sur des algorithmes éprouvés comme l'algorithme EM (*expectation-maximisation*) [2]. La méthode de détermination des séquences se distingue du premier courant en ce sens que toutes les séquences possibles, présentes dans le corpus d'apprentissage, sont présentées à l'initialisation de l'algorithme. Ce n'est qu'au cours des itérations successives que les séquences inutiles sont peu à peu supprimées. L'algorithme d'apprentissage d'un modèle unigramme de multi-

grammes opère comme suit :

1. Énumérez l'ensemble des séquences présentes dans le corpus d'apprentissage à concurrence d'une longueur maximale. Donner à chaque séquence ou mot isolé sa fréquence dans le corpus comme probabilité unigramme ;
2. Supprimer les séquences les plus rares selon un seuil fixé *a priori* ;
3. Ré-estimer ces probabilités à l'aide d'un algorithme *forward-backward* utilisant le modèle unigramme courant ;
4. Supprimer de l'ensemble des séquences, les plus rares (utilisation d'un seuil fixé *a priori*) selon le nouveau modèle unigramme ;
5. Itérer jusqu'à convergence de la perplexité.

Nous reportons le lecteur aux travaux de S. Deligne [1] pour une formalisation des multigrammes ainsi que des algorithmes *ad hoc*. L'algorithme *forward-backward* permet de sommer, de manière efficace, les vraisemblances de toutes les segmentations possibles du corpus d'apprentissage d'origine en utilisant les séquences du vocabulaire. Le principe des multigrammes propose à la fois une méthode d'apprentissage d'un jeu optimal de séquences à partir d'un jeu initial, ainsi qu'un algorithme permettant de calculer la vraisemblance d'une suite de mots en utilisant les séquences ainsi obtenues. Il est donc adapté à notre problème qui est, rappelons le, d'intégrer un jeu de séquences toutes déterminées *a priori*. Toutefois, la méthode, du fait de sa combinatoire initiale importante, n'est pas adaptée directement aux grands vocabulaires. En effet, dans ce cas, le nombre de séquences possibles à l'initialisation devient beaucoup trop important. Nous avons donc adapté le principe des multigrammes en proposant à l'initialisation de l'algorithme d'apprentissage non l'ensemble des séquences possibles, mais un jeu restreint obtenu à la suite de nos travaux sur le Principe de Sélection par l'Historique (SHP).

3. Construction d'un jeu de séquences selon le principe SHP

Une séquence comme « pomme de terre » peut être considérée comme étant construite par la liaison des deux mots « pomme » et « terre » à l'aide du mot outil « de ». Cette construction donne naissance à une unité dont le sens n'est pas une fonction simple des sens de ses composants. Dans cette relation entre les deux mots « pomme » et « terre », les deux mots sont distants l'un de l'autre. On peut penser de même à de nombreux autres exemples mettant en jeu des relations à des distances différentes comme, par exemple, « metteur en scène », « président de la République », « pince à linge »... En revanche, toute occurrence d'une telle construction ne correspond pas à une unité linguistique. Ainsi, dans le groupe nominal « le livre de Kamel », « livre de Kamel » ne peut prétendre au statut de séquence.

3.1. Modèles distants

Pour déterminer les suites de mots qui peuvent devenir des séquences dans le cadre d'un modèle statistique de langage, nous utilisons des modèles de bigrammes distants [6] avec des distances différentes. A chaque valeur de la distance d est associé un modèle distinct défini par :

$$\begin{aligned} P_d(w_i|w_1 \dots w_{i-1}) &\overset{\text{approx}}{=} P_d(w_i|w_{i-d-1}) \\ &\overset{\text{def}}{=} \frac{N_d(w_{i-d-1}, w_i)}{N(w_{i-d-1})} \end{aligned} \quad (1)$$

où $N_d(v, w)$ est le nombre de fois que v et w ont été rencontrés dans le corpus d'apprentissage séparés par d mots. On notera que quand d est nul, on revient au cas du modèle bigramme.

Pour déterminer que la suite de mots « président de la République » peut être candidate au statut de séquence, on part de l'idée de **relation intra-lexicale distante** que nous illustrons par l'exemple suivant : *Si le modèle de bigrammes distants de distance 2 permet de prédire avec « beaucoup » plus de certitude le mot suivant « président de la » que ne le peuvent les modèles de bigrammes distants de distance 1 et 0, alors le lien entre « président » et le mot suivant « président de la » est suffisamment fort pour conclure à l'existence d'une séquence.*

Il reste alors à définir une fonction qui mesure le pouvoir prédicteur de chacun des modèles distants ou non les uns par rapport aux autres, et ce pour un même historique donné.

3.2. Mesure du pouvoir prédicteur d'un modèle statistique de langage

La perplexité est une mesure communément utilisée pour évaluer un modèle statistique de langage. Elle est définie par la formule :

$$PP = \left(\prod_{i=1}^{N} P(w_i|h_i) \right)^{-\frac{1}{N}} \quad (2)$$

où N est la taille du corpus utilisé pour le calcul. Cette mesure représente le facteur de branchement moyen du modèle de langage [7], et donc sa capacité moyenne de prédiction : plus ce facteur est faible, plus sûre est la prédiction d'un mot à la suite d'un historique. La perplexité est une moyenne calculée sur tous les événements contenus dans le corpus utilisé. Ainsi, elle confond les capacités prédictives du modèle en une seule mesure, indépendante de l'historique. Or, la capacité de prédiction d'un modèle de langage est très dépendante de l'historique. On se propose donc dans la suite de ré-écrire la formule de la perplexité afin d'extraire une mesure de cette capacité pour un historique donné.

Souvent, on utilise la formule (2) sous sa forme logarithmique afin de manipuler une somme :

$$\log PP = -\frac{1}{N} \sum_{i=1}^{N} \log P(w_i|h_i) \quad (3)$$

On peut ré-organiser cette somme en regroupant les occurrences des mêmes événements. Ici, un événement est un historique suivi d'un mot :

$$\log PP = -\frac{1}{N} \sum_{hw,\ hw \in \mathcal{C}} N(hw) \log P(w|h) \quad (4)$$

où $N(hw)$ est le nombre de fois que la suite hw est rencontrée dans le corpus utilisé, ici noté $\mathcal{C}$. On peut maintenant factoriser les termes de même historique et obtenir :

$$\log PP = -\frac{1}{N} \sum_{h \in \mathcal{C}} \left(\sum_{w_i,\ hw_i \in \mathcal{C}} N(hw_i) \log P(w_i|h) \right) \quad (5)$$

En définissant $Q(h) = \sum_{w_i,\ hw_i \in \mathcal{C}} N(hw_i) \log P(w_i|h)$, on obtient finalement :

$$\log PP = -\frac{1}{N} \sum_{h \in \mathcal{C}} Q(h) \quad (6)$$

Ainsi définie, $Q(h)$ est constituée des termes de la perplexité correspondant à l'historique h, et seulement à cet historique. Comme la perplexité est une mesure moyenne de la capacité prédictive du modèle sur l'ensemble des historiques, $Q(h)$ définie comme la restriction de PP à h, peut être considérée comme une mesure de cette capacité pour l'historique h. Cette mesure a été utilisée en premier lieu pour une combinaison entre modèles fondées sur la sélection d'un modèle parmi d'autres. Pour ce faire, le modèle de langage choisi, pour un historique donné h, est celui qui obtient la valeur $Q(h)$ la plus haute. Ceci permet de combiner des modèles plus efficacement que ne le permet une combinaison linéaire [9]. Comme le principe consiste à choisir le modèle en fonction de l'historique, nous appelons cette approche le principe de Sélection par l'Historique (SHP).

3.3. Choix des candidats au statut de séquence

La mesure définie ci-dessus peut être utilisée pour comparer le pouvoir prédicteur de plusieurs modèles de langage entre-eux, pour un même historique donné. On peut ainsi comparer un modèle de langage distant à un modèle non distant. Partant de l'idée de relation intra-lexicale distante, un historique h tel que le modèle de langage distant aurait un pouvoir prédicteur beaucoup plus important que celui du modèle non distant peut alors être utilisé pour former des suites de mots candidates au statut de séquence.

Nous avons ainsi comparé un modèle de bigrammes distants de distance 1 par rapport à un modèle bigramme (distance nulle) d'une part, et un modèle de bigrammes distants de distance 2 par rapport au même modèle bigramme, d'autre part. Le premier jeu de comparaisons permet de construire des séquences de longueur 3, alors que le deuxième donne naissance à des séquences de longueur 4. Nous donnons ci-dessous la méthode utilisée pour construire les séquences de longueur 3 :

1. Utiliser un corpus dit d'extraction $\mathcal{C}$;
2. Pour chaque historique $w_{i-2}w_{i-1}$ d'occurrence supérieure ou égale à α dans $\mathcal{C}$:
   – Si $Q_1(w_{i-2}w_{i-1}) > Q(w_{i-2}w_{i-1})$ alors inclure toutes les suites $w_{i-2}w_{i-1}w_i$ de $\mathcal{C}$ dans l'ensemble des candidates au statut de séquence.

$Q_1(w_{i-2}w_{i-1})$ est la capacité prédictive du modèle de bigrammes distants (distance égale à 1) pour l'historique $w_{i-2}w_{i-1}$ et $Q(w_{i-2}w_{i-1})$ celle du modèle de bigrammes non distants. Nous avons choisi expérimentalement la valeur 10 pour seuil α. Ce seuil permet de ne pas comparer deux modèles de langage sur des historiques trop rares et donc des données trop peu nombreuses.

Nous avons appliqué cette méthode à la détection de séquences de longueur 3 et 4. Bien que les données d'apprentissage et d'extraction utilisées ne sont décrites que par la suite, nous donnons pour information, dans la table 1, un échantillon représentatif de ces séquences parmi les meilleures en terme d'écart entre $Q_1(w_{i-2}w_{i-1})$ et $Q(w_{i-2}w_{i-1})$. On remarque que les séquences ainsi déterminées sont intuitivement représentatives de la langue. Les séquences ont ceci de particulier qu'elles correspondent aussi bien à des unités lexicales (« tremblement de terre ») qu'à des expressions toutes faites comme « joindre les deux bouts », de type verbale ou nominale, qui font toute l'identité d'une langue.

4. CONSTRUCTION ET ÉVALUATION DE MODÈLE DE SÉQUENCES

Nous avons estimé les modèles de langage en utilisant un vocabulaire de 20K mots et 38M de mots d'apprentissage issus du corpus journalistique *Le Monde* (années 87-88). Tous les paramètres absents du corpus d'apprentissage ont été estimés en utilisant le principe du repli (*backing-off*)

TAB. 1: Échantillon des « meilleures » suites de trois et quatre mots ayant une forte relation intra-lexicale distante aux extrémités.

Suites de 3 mots	Suites de 4 mots
voler en éclats	pain sur la planche
vice de procédure	joindre les deux bouts
vents et marées	inventaire à la Prévert
valet de chambre	interruption volontaire de grossesse
vaille que vaille	infraction à la loi
vache à lait	impôt sur le revenu
tueurs à gages	feux de la rampe
tremblement de terre	course contre la montre
tirage au sort	couler beaucoup d' encre
seigneur tout honneur	colle à la peau

TAB. 2: Performances des modèles sans séquences en terme de perplexité.

Modèle	Perplexité
bigramme	127.1
trigramme	89.1

avec la méthode *absolute-discounting* [3]. Le corpus d'extraction utilisé afin de déterminer les séquences est constitué de 10M de mots provenant de trois corpus, tous distincts du corpus d'apprentissage : 2M de mots du journal *Le Monde*, 6M provenant du journal *Le Monde Diplomatique* et 2M provenant de la version en ligne du journal Algérien de langue Française *El Watan*. Enfin, le corpus de test utilisé est constitué de 2M de mots issus du corpus *Le Monde*. Ce dernier corpus n'a aucune partie commune avec les autres corpus.

4.1. Modèles de référence

Les performances des modèles bigramme et trigramme en terme de perplexité ont été estimées sur ce corpus de test. Ces performances sont données dans la table 2.

4.2. Modèle de séquences

Nous utilisons l'algorithme adapté d'apprentissage des multigrammes en prenant comme jeu initial de séquences les séquences de longueur 3 et 4 telles qu'obtenues par la méthode SHP. De plus, conscients de l'importance des séquences de longueur 2 [4], nous intégrons aussi tous les couples de mots présents dans le corpus d'apprentissage. De toutes ces séquences candidates, nous rejetons les séquences présentes moins de 10 fois dans le corpus d'apprentissage. Finalement, en réunissant les ensembles de suites de 2, 3 et 4 mots, nous avons à disposition environ 3.5M d'unités pouvant être ajoutées aux 20K formes de base. Parmi ces séquences, 1.4M sont de longueur 3 et 120K séquences sont de longueur 4. Comme condition de convergence de l'algorithme, nous avons dans un premier temps choisi la perplexité sur le corpus de développement de 2M de mots issus du journal *Le Monde*. A chaque itération, nous supprimons de l'ensemble des séquences les plus rares (10% de la totalité des séquences) selon le modèle de séquences estimé à cette étape. Nous avons opéré l'apprentissage d'un modèle bigramme de séquences en utilisant le principe des multigrammes adapté aux grands vocabulaires. Le vocabulaire optimal obtenu en terme de perplexité est de taille 40K dont 20K séquences réparties comme suit : 18.2K séquences de longueur 2, 1.6K de longueur 3 et 0.2K de longueur 4. Ces effectifs confirment, comme nous l'avions prévu ci-dessus, l'importance des séquences de longueur 2. Le modèle bigramme de séquences ainsi construit obtient une perplexité de 100.2 sur le corpus de test, soit une amélioration de 21.2% par rapport à la perplexité du modèle bigramme de mots simples (127.1).

La taille du vocabulaire ainsi obtenu est le double de celle du vocabulaire initial. Or, les séquences sont destinées à être intégrées dans un système de reconnaissance. C'est pourquoi, nous avons décidé d'utiliser aussi un vocabulaire de séquences de taille plus restreinte. Pour cela, nous avons

laissé l'algorithme se dérouler jusqu'à atteindre un nombre de séquences raisonnable. Nous avons fixé cette taille à 4000, suite à des travaux précédents [11].

5. APPLICATION À LA RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

Suite aux expérimentations précédentes, nous avons obtenu deux modèles bigrammes de séquences, l'un utilisant 20K séquences, et l'autre 4000 séquences. Nous avons intégré ces modèles de séquences dans le système Sirocco. Sirocco [5] (`http ://www.irisa.fr/sirocco/`) est un système fondé sur l'algorithme de reconnaissance proposé par S. Ortmanns et H. Ney [10]. Le système fonctionne en deux passes. La première génère un graphe de mots dirigé et acyclique. Chaque transition est décrite par un score acoustique et un score linguistique. Cette première passe utilise un modèle bigramme. La deuxième permet d'utiliser un modèle plus performant, comme un modèle trigramme, pour rechercher dans le graphe la suite de mots ayant le meilleur compromis entre vraisemblance acoustique et vraisemblance linguistique. L'unité acoustique utilisée est le phonème. A chacun est associé un Modèle de Markov Caché à trois états connectés de gauche à droite sans saut possible du premier au troisième état. Le signal acoustique a été échantillonné à 16kHz. Les vecteurs acoustiques sont formés de 35 coefficients cepstraux : énergie (11 coefficients), premières et deuxièmes dérivées (12 coefficients pour chaque liste). Pour la transformation cepstrale, une fenêtre de largeur 512 échantillons et un décalage de 128 ont été utilisés. A chaque état de HMM est associée une distribution de probabilités, définie par une mixture de 16 gaussiennes de matrice de covariance diagonale. Les données de BREF80 [8] ont été utilisées pour l'apprentissage des modèles acoustiques. Pour l'ensemble des 20K formes graphiques de base utilisées par le système, on manipule 64K prononciations différentes. Les prononciations possibles d'une séquence ont été générées par énumération des concaténations des prononciations des mots composant cette séquence.

Nous comparons le système utilisant les séquences à la version de base. Celle-ci utilise en première passe un modèle bigramme de mots simples et analyse les graphes avec un modèle trigramme de mots simples. Les versions utilisant les séquences utilisent en première et deuxième passe un modèle bigramme de séquences. Les trois versions de Sirocco ont été testées sur 300 phrases non utilisées pour l'apprentissage des modèles acoustiques. Les performances de ces deux versions en terme de mots reconnus, substitués, supprimés et insérés, ainsi qu'en terme de *Word Error Rate*, sont présentées dans la table 3. On remarque que l'adjonction de 4000 séquences permet d'augmenter le taux de mots reconnus de 8.7%, et de diminuer le taux de substitution et de suppression respectivement de 15.2% et de 9.7%. Le taux d'erreur global, quant à lui, est diminué de 12.7%. Il semble que l'utilisation de 20K séquences donne des performances légèrement inférieures. Ceci peut s'expliquer par le fait que cela revient à utiliser un trop grand nombre de prononciations pour un même vocabulaire de base. Une première analyse qualitative des résultats montre que les séquences de toutes longueurs (2, 3 et 4) sont utiles en terme de reconnaissance. De plus, en général, quand une séquence est bien reconnue, cela a un impact positif en reconnaissance sur son voisinage dans la phrase dictée. Un autre point est que, souvent, une séquence proposée par le système, même si elle ne correspond pas aux mots dictés, est pourtant acoustiquement plus proche de la portion du signal correspondante que la suite de mots proposée par le système de référence. Enfin, on note que les séquences souvent utilisées sont celles qui contiennent un mot outil. Ces mots, généralement courts sont difficiles à reconnaître, car sujets à des déformations du fait des contraintes acousticoarticulatoires. Les insérer dans des séquences semble donc permettre de renforcer leur robustesse en reconnaissance.

TAB. 3: Performances du système Sirocco selon que les séquences sont utilisées ou non en première passe, en terme de mots reconnus (REC), substitués (SUBS), supprimés (SUPP), insérés (INS), et de taux d'erreur (WER).

Version	REC	SUBS	SUPP	INS	WER
sans séquences	61.9	30.9	7.2	4.5	42.6
avec séquences (20K)	66.8	27.0	6.3	4.5	37.8
avec séquences (4K)	67.3	26.2	6.5	4.5	37.2

6. CONCLUSION

Nous avons présenté dans cet article le principe de Sélection par l'Historique, permettant de construire un jeu de séquences candidates à l'introduction dans le vocabulaire du modèle bigramme d'un système de RAP. Le principe multigramme a permis alors de sélectionner les séquences les plus utiles en terme de perplexité. Finalement, les séquences obtenues ont permis d'améliorer le modèle bigramme de base de 21% en Perplexité et le taux de reconnaissance du système de 8.7%. Le taux d'erreur a été diminué de 12.7%. Les séquences obtenues semblent avoir une validité linguistique certaine. Toutefois une étude reste à mener afin de confirmer leur utilité d'un point de vue linguistique. De plus, la même démarche, telle que présentée dans cet article, peut être utilisée avec d'autres heuristiques de détermination d'un jeu initial de séquences. On peut penser, par exemple, à utiliser le constat sur la robustesse des mots courts en reconnaissance quand ils sont intégrés dans des séquences, et systématiser leur intégration. Enfin, un point important est aussi de mettre en œuvre un modèle de trigrammes de séquences.

RÉFÉRENCES

[1] S. Deligne. *Modèles de séquences de longueur variable : application au traitement du langage écrit et de la parole*. PhD thesis, Télécom Paris, 1996.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistics Society*, 39(1) :1–38, 1977.

[3] M. Federico and R. De Mori. *Spoken dialogues with computers*, chapter Language Modelling, pages 199–230. Academic Press, 1997.

[4] E. Giachin. Phrase bigrams for continuous speech recognition. In *Proc. of ICASSP*, pages 225–228, 1995.

[5] G. Gravier and F. Yvon. Notes on the Sirocco project. (accés web : http ://www.enst.fr/ sirocco/public/sirocco.ps), 2001.

[6] X. Huang, F. Alleva, H.-W. Hon, M.-Y. Hwang, K.-F. Lee, and R. Rosenfeld. The SPHINX speech recognition system : an overview. *Computer Speech and Language*, 2 :137–148, 1993.

[7] Frederick Jelinek. Self-organized language modeling for speech recognition. In A. Waibel and K.-F. Lee, editors, *Readings in Speech Recognition*, pages 450–506. Kaufmann Publishers, San Mateo, CA, 1990.

[8] L. Lamel, J.-L. Gauvain, and M. Eskenazi. BREF, a large vocabulary spoken corpus for french. In *Proc. of Eurospeech*, volume 2, 1991.

[9] D. Langlois, K. Smaïli, and J.-P. Haton. Efficient language models combination : Application to phrase finding. In *Proceedings of the International Workshop "Speech and Computer"*, 2001.

[10] S. Ortmanns and H. Ney. A word graph algorithm for large vocabulary continuous speech recognition. *Computer Speech and Language*, 11 :42–72, 1997.

[11] I. Zitouni. *Modélisation du langage pour les systèmes de reconnaissance de la parole destinés aux grands vocabulaires : application à MAUD*. PhD thesis, 2000.