Décodage acoustico-phonétique dans un système de Reconnaissance Automatique de la Parole Grand Vocabulaire

(Dictée Vocale)

Application dans le projet SIROCCO

Bruno Jacob

Séminaire AMETIST-CORDIAL, Lannion, 29-30 mars 2000

Modélisation de l'espace de recherche

But : Retranscrire un signal acoustique en séquences de mots (phrases)

→ On cherche une structure qui modélise les phrases possibles:

1 seul modèle global est impossible à réaliser
Donc, on "découpe" cette structure en niveaux de modélisation.

Généralement, on ne sait modéliser que 3 niveaux :

acoustique : modélisation des unités sonores

$\updownarrow$ articulation des sons en mots

lexical : modélisation des mots

$\updownarrow$ articulation des mots en phrases

syntaxique : modélisation des phrases

Cadre de travail pour le décodage

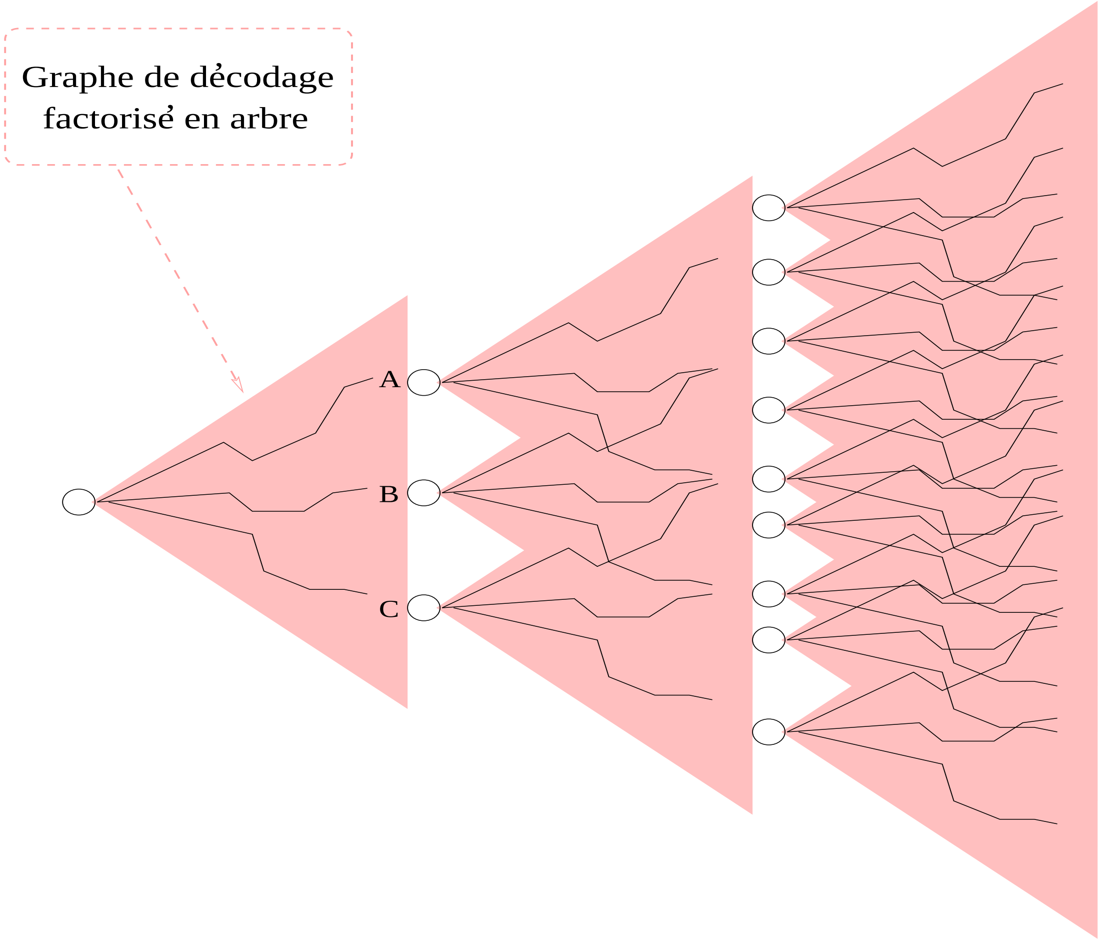
Dans les systèmes de reconnaissance performants le décodage se fait **simultanément** dans les 3 niveaux de modélisation ($\neq$ avant $\rightarrow$ successivement dans les niveaux $\Rightarrow$ propagation d'erreurs)

Le module de décodage

- pilote conjointement:
  - le *Graphe de Décodage* $\rightarrow$ informations aux niveaux lexical et acoustique
  - le *Modèle de Langage* $\rightarrow$ informations au niveau syntaxique
- organise l'espace de recherche des phrases possibles.

Exploitation d'un graphe de décodage

Exemple avec un vocabulaire $\{A, B, C\}$



Décodage acoustico-phonétique dans Sirocco

(Exploitation du graphe de décodage)

Difficultés :

- taille du lexique > 100.000 mots ($\Rightarrow$ Énorme graphe)
- complexité du Modèle de Langage

$\Rightarrow$ Recherche classique d'une phrase impossible dans ce contexte d'un système de Dictée Vocale

Méthodes :

- Recherche en **faisceaux** par l'algorithme de **Viterbi** (*Beam-Search*)
- Contrôle de la largeur des faisceaux $\Rightarrow$ taille espace de recherche

Principe

⇒ Décorrélation du calcul des hypothèses intra-mots et inter-mots :

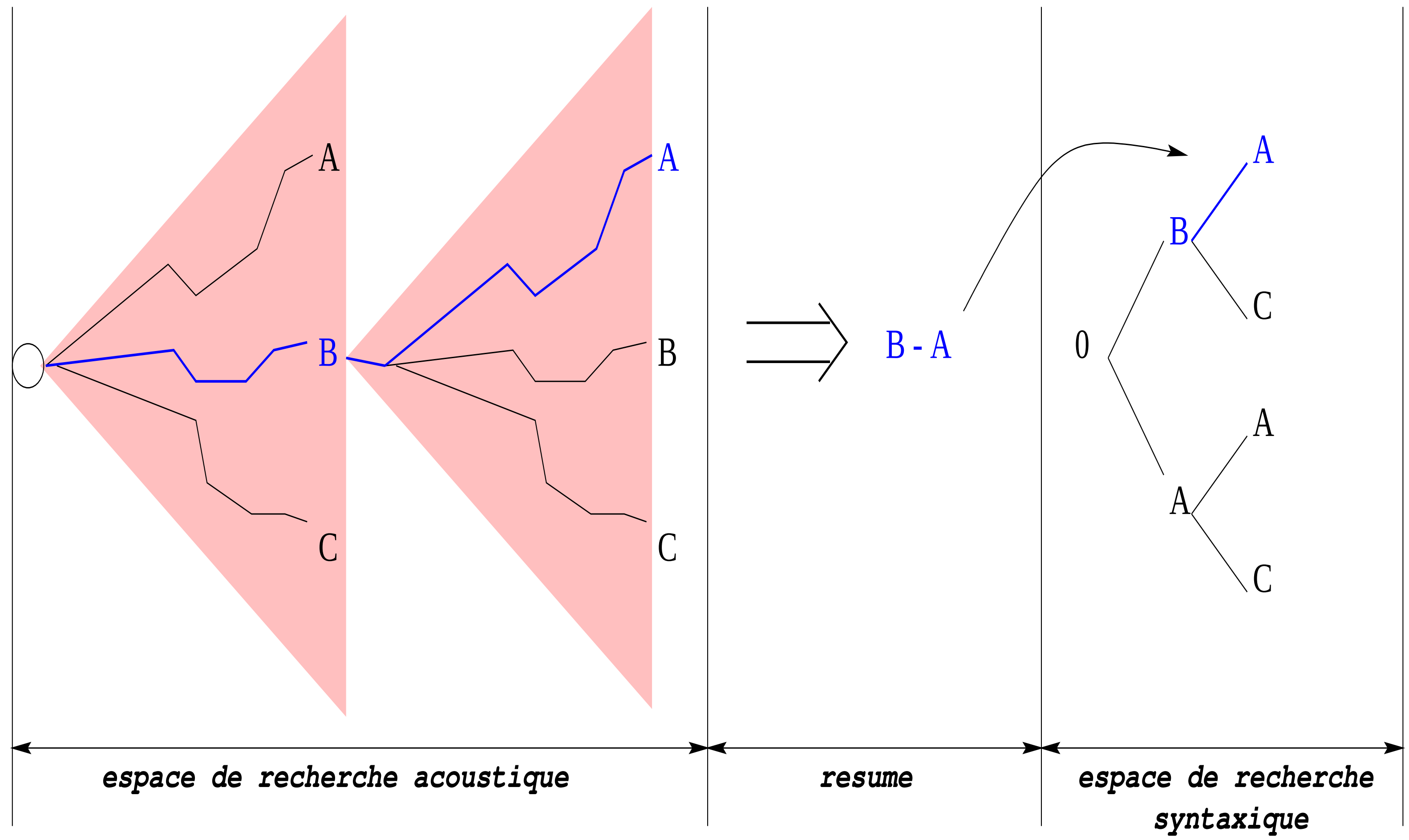
- Développement de la "totalité" de l'espace de recherche à l'intérieur des mots (niveau acoustique)

  `Ambiguïté de la recherche acoustique élevée → l'effort sur le décodage acoustique doit être important`

- L'espace de recherche inter-mots ne manipule que des "résumés" des solutions intra-mots

  `On peut ainsi utiliser des ML complexes et de tailles importantes sans avoir besoin d'un espace mémoire prohibitif`

Exemple d'utilisation avec un Modèle de Langage bigramme :



Contrôle de la taille de l'espace de recherche

Au niveau intra-mots : Élagage acoustique

- *standard* : % du score maximum au temps t
- *histogramme* : nombre max d'hypothèses au temps t
- *look-ahead* : recherche "en avant" dans le début de décodage des mots

Au niveau inter-mots : Élagage

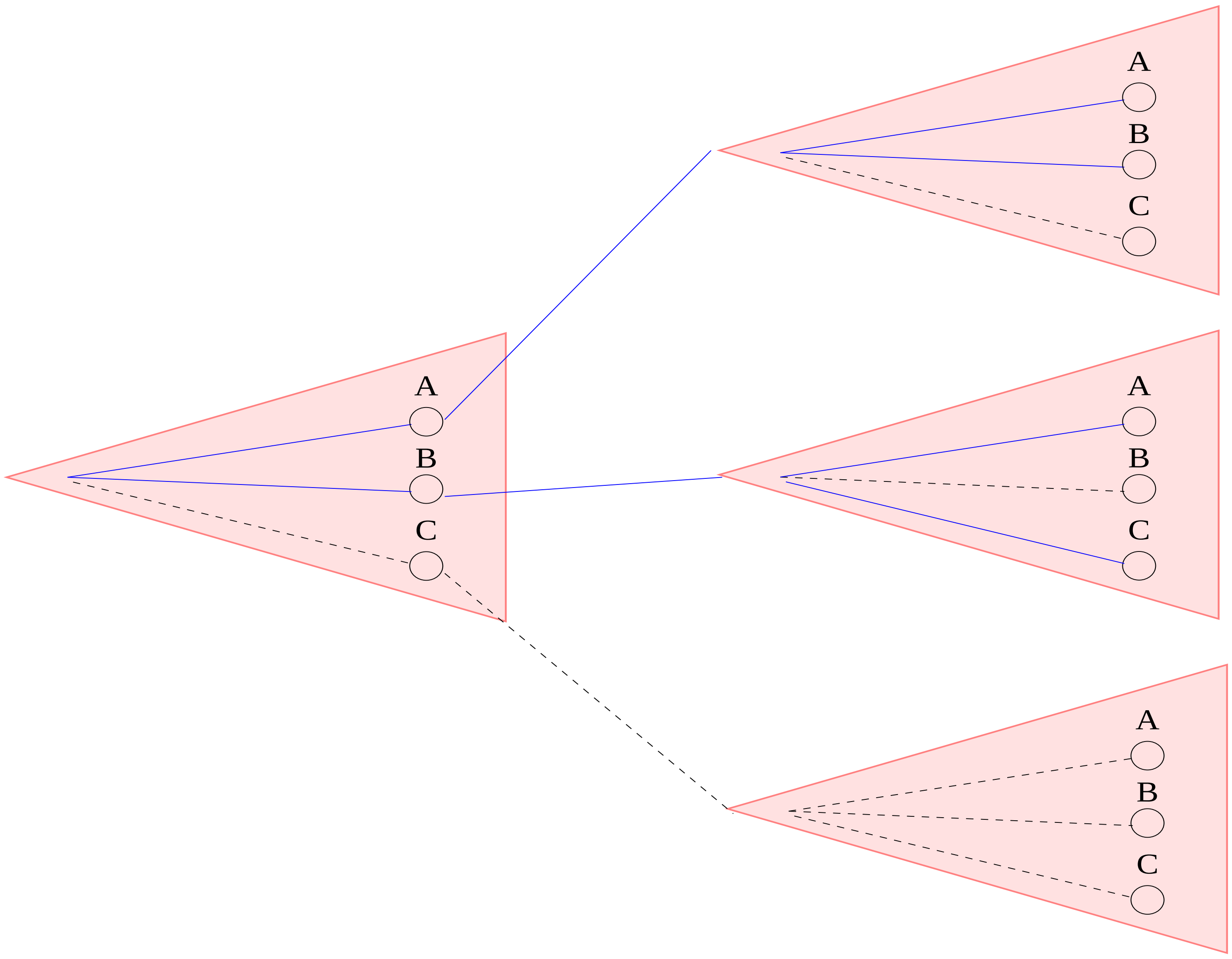
- *standard* : % du score maximum des mot terminés au temps t
- *histogramme* : nombre max de mots par trame t
- *purge* : les chemins perdus

$\Rightarrow$ détermine la largeur du faisceau de recherche (*Beam-Search*)

Avantages :

- gains d'espace mémoire et vitesse d'exécution

Faisceaux dans l'espace de recherche acoustique pour un ML bigramme



1^{er} test du système sur un vocabulaire réduit:

- Lettres épelées du français + mots clés ($\approx$ 37 mots)
- 37 modèles de phonèmes + 3 modèles de silence
- 5 états (3 émetteurs)

Test sur un fichier de 1069 trames = environ 11 secondes

- recherche complète : $>$ 24 heures
- recherche en faisceaux :

Max hypothèses acoustiques à chaque trame	Temps d'exécution (environ)	Temps Réel
1000	30"	$\times$ 2,5
5000	2'	$\times$ 12
10000	5'	$\times$ 27
50000	30'	$\times$ 150

Conclusion

Validation avec HTK :

phrase de référence	%Corr SIROCCO	%Corr HTK
e r v é	50.00	25.00
m a r i a trait d'union l u i s a	58.33	58.33
i a c o b a c deux c i o	60.00	40.00
a z u c e n a	28.57	28.57
f o o d l i n e	0.00	0.00

→ 1^{ere} étape de la validation réussie

⇒ Test du système Sirocco sur un vocabulaire de 20.000 mots