

Modèles de langage à portée variable : Application au traitement des homophones

Frédéric Béchet, Alexis Nasr, Thierry Spriet, Renato de Mori

Laboratoire Informatique d'Avignon – Université d'Avignon
Agroparc BP 1228 – 84911 Avignon Cedex 9
frederic.bechet@lia.univ-avignon.fr <http://www.lia.univ-avignon.fr/>

Résumé

L'objectif de cette étude concerne le traitement d'homophones singulier/pluriel dans un Système de Reconnaissance de la Parole en exploitant les contraintes d'accord dans la phrase à reconnaître. Un certain nombre de ces contraintes ne peut être traité par les modèles de langage à portée locale de type n-gram utilisés habituellement. Les deux modèles proposés, le modèle à base de syntagme et le modèle Homophone-Cache, permettent de résoudre certains cas d'homophonie par deux méthodes différentes : le modèle à base de syntagme permet d'introduire des contraintes syntaxiques ; le modèle Homophone-Cache a pour objet de discriminer les homophones singulier/pluriel, de manière robuste, en étant peu sensible à la mauvaise reconnaissance d'un mot au sein de la phrase.

1. Introduction

Les modèles de langage utilisés dans les Systèmes de Reconnaissance Automatique de la Parole (SRAP) ont pour but de garantir la cohérence syntaxique et sémantique des phrases produites par le système. C'est en particulier les modèles de langage qui permettent de choisir entre plusieurs formes homophones d'un même mot, ce qui constitue un des problèmes cruciaux des SRAP appliqués à la langue française.

La langue française se caractérise en effet par son fort taux d'homophonie. Une étude d'un lexique phonétisé de 240k mots montre que le nombre moyen d'homophones par mot est de 1.8. De plus, ce phénomène d'homophonie est fortement lié à des considérations morphologiques : si l'on considère la flexion en nombre des 40.7k lemmes composant le lexique, nous constatons que 72% d'entre eux ont des formes fléchies singulier/pluriel qui sont homophones.

Dans un SRAP, lors de la phase de décodage, seul le modèle de langage peut lever l'ambiguïté de transcription sur un mot homophone singulier/pluriel. Si la plupart de ces ambiguïtés peuvent être levées de manière locale (dans les séquences <déterminant> <nom>, par exemple), une analyse syntaxique plus poussée est nécessaire pour prendre en compte des phénomènes d'accord plus distants.

Cette analyse syntaxique est elle-même inefficace dans certains cas où seules des informations sémantiques ou lexicales permettent de déterminer le nombre d'un lemme (par exemple : "il y a plus de pommes cette année" et "il y a plus de vent cette année").

Ce travail s'inscrit dans les nombreuses études menées actuellement pour proposer des modèles de langage à portée plus grande que les modèles 3-gram classiques (Bellegarda 1998; Deligne 1998; Zitouni 1998).

Nous présentons dans cette étude une combinaison de modèles de langage permettant de prendre en compte des contraintes d'accord s'étendant sur l'ensemble de la phrase. Ces modèles permettent de traiter des graphes d'homophones en se focalisant sur le problème des mots homophones singulier/pluriel.

2. Génération de graphes d'homophones

Un graphe d'homophone est obtenu en remplaçant chaque mot d'une phrase par l'ensemble de ses homophones. En considérant tous les chemins possibles on obtient un graphe dans lequel on peut tester différents modèles de langage. Ainsi, l'évaluation se fait en mesurant le nombre d'ambiguïtés correctement levées. Cette technique est une alternative intéressante à l'évaluation intrinsèque des modèles de langage par la seule perplexité. De plus, elle nous permet de tester le modèle sur de très grands corpus de test.

En utilisant un dictionnaire de 240K mots phonétisés par notre système de transcription graphème-phonème (Béchet 97), nous sommes en mesure de produire de tels graphes comme dans l'exemple ci-dessous (Table 1) où les homophones en nombre d'un même mot sont représentés dans des cases grisées.

<i>il</i>	<i>a</i>	<i>pris</i>	<i>la</i>	<i>parole</i>	<i>en</i>	<i>public</i>
hile	A	prie	la	parole	an	public
hiles	a	prient	là	paroles	ans	publics
il	à	pries	las		en	publique
ils	ah	pris			han	publiques
île	ha	prît				
îles		prix				

Table 1 : graphe d'homophones généré à partir d'une phrase

Nous nous sommes focalisés dans cette étude sur les lemmes qui ont leurs formes singulier et pluriel homophones, les résultats présentés ont été obtenus sur des graphes ne contenant que ce type d'homophones.

3. Fusion de modèles

La stratégie de décodage des graphes d'homophones utilisée dans cette étude est de type "stack-decoding". Des probabilités sont associées à chaque chemin possible dans le graphe et la phrase solution est le chemin qui obtient la meilleure probabilité d'après nos modèles de langage.

Le modèle de langage que nous proposons est une combinaison de plusieurs modèles : un modèle 3-gram (M1) sur les graphies, un modèle 3-class (M2) sur des catégories syntaxiques associées aux mots, un modèle 3-gram (M3) sur les syntagmes composant la phrase et, enfin, un modèle spécifique au traitement des homophones singulier/pluriel que nous appelons

Modèles de langage à portée variable

modèle Homophone-Cache (M4). La probabilité d'une phrase représente la combinaison linéaire des probabilités de chacun des modèles sur la phrase.

Soit une phrase W , la probabilité $P(W)$ associée aux mots de W par l'ensemble des modèles est égal à :

$$P(W) = \alpha_1 P_{M1}(W) + \alpha_2 P_{M2}(W) + \alpha_3 P_{M3}(W)$$

Le modèle M4 est utilisé en amont, comme nous le verrons dans le paragraphe présentant le modèle Homophone-Cache.

3.1. Modèle n -gram

Les modèles de langage utilisés habituellement en RAP sur de grands vocabulaires utilisent une approche statistique de type n -gram (2 ou 3-gram) permettant de prendre en compte des contraintes locales d'accord. Ainsi, avec un modèle 3-gram, la probabilité d'un chemin W contenant n mots m est :

$$P_{M1}(W) = \prod_{i=1}^n P(m_i | m_{i-2}, m_{i-1})$$

Cependant, même pour le cas de contraintes locales, il suffit que le n -gram n'ait pas été vu lors de l'apprentissage pour que le repli vers des modèles d'ordre inférieur empêche toute cohérence d'accord.

Nous voyons dans l'exemple ci-dessous (Table 2) l'effet d'un mauvais apprentissage du modèle : le 2-gram [*salaires*, *évoluent*] n'a pas été vu lors de l'apprentissage, le repli du modèle vers la probabilité du 1-gram [*évolue* / *évoluent*] va conduire au choix de la forme singulier plus fréquente dans le corpus d'apprentissage du modèle.

Graphe	les	salaires	évoluent	de	façon	mesurée
Décodage	les	salaires	évolue	de	façon	mesurée

Table 2 – décodage d'un graphe d'homophone par un modèle 3-gram

Nous utilisons un modèle 3-gram avec repli sur 2gram et 1-gram, basé sur un lexique de 64K mots, et entraîné sur des textes issus du journal Le Monde entre 1987 et 1993 (environ 100M mots).

3.2. Modèle n -classes

L'utilisation de classes syntaxiques associées aux mots permet de traiter efficacement le problème de l'absence d'un n -gram de mots en considérant la séquence n -gram de classes. Nous utilisons un jeu de 105 classes syntaxiques et un étiqueteur statistique (Spriet, 95) développé en interne. Le corpus d'apprentissage de notre modèle 3-classe est constitué des mêmes textes issus du journal Le Monde, étiquetés par notre système.

Dans ce modèle, la probabilité d'un chemin W composé de n mots m auxquels sont associés des classes syntaxiques c est :

$$P_{M2}(W) = \prod_{i=1}^n P(m_i | c_i) \times P(c_i | c_{i-2}, c_{i-1})$$

L'exemple précédent est parfaitement traité par notre modèle 3-classe. Néanmoins ces modèles se révèlent inefficaces dès que les contraintes d'accord s'étendent au delà de l'historique concerné.

Par exemple, la phrase suivante (Table 3) sera mal décodée par notre modèle 3-classe en raison du groupe prépositionnel rattaché au sujet.

<i>Mots</i>	<i>Catégorie syntaxique</i>	<i>Décodage</i>
Les	DETMP	Les
cours	NMP	cours
de	PREP	de
la	DETFS	la
bourse	NFS	bourse
continue <i>ou</i> continuent	V3S <i>ou</i> V3P	continue
de	PREP	de
s'	PREFMS	s'
effondrer	VINF	effondrer

Table 3 – décodage d'un graphe d'homophone par un modèle 3-classe

3.3. *Modèle Syntagme*

La prise en compte de phénomènes d'accord entre mots distants, tel que dans la phrase d'exemple précédente, nécessite des modèles de langage ayant une portée plus grande que l'historique communément utilisé de 2 mots. Le manque de données d'apprentissage interdit généralement le recours à des n-gram d'ordre supérieur. Une solution possible consiste alors à regrouper plusieurs mots au sein d'unités d'ordre supérieur (des syntagmes), provoquant ainsi un "rapprochement" des syntagmes concernés par un accord.

Différentes approches ont été utilisées pour obtenir ces syntagmes de tailles variables, soit à base de grammaire stochastiques (Gillett 1998), soit à base de critères purement statistiques (Deligne 1995). Notre approche relève de ces deux types de méthodes, elle utilise à la fois des règles de réécriture syntaxiques permettant de reconnaître certains syntagmes et d'autre part un modèle stochastique représentant les probabilités d'occurrence d'un syntagme d'un type et d'un nombre donné en fonction des deux syntagmes le précédant.

Le corpus utilisé dans la phase d'apprentissage du modèle est composé d'articles du journal Le Monde et comprend environ 100M de mots.

Lors de la phase d'apprentissage, ce corpus est tout d'abord étiqueté par l'étiqueteur stochastique. Puis, il est partiellement analysé à l'aide d'un analyseur syntaxique partiel à états finis utilisant les classes précédemment posées. Le but de cette analyse est de regrouper des suites de classes constituant un syntagme nominal, prépositionnel ou verbal en gardant trace du nombre du syntagme lorsque cette information est pertinente.

La couverture de la grammaire est volontairement limitée, ceci pour deux raisons. D'une part, l'objectif n'est pas d'analyser entièrement une phrase, mais uniquement certains syntagmes entre lesquels existent des phénomènes d'accord. D'autre part, la couverture réduite nous permet d'ignorer la plupart des cas d'ambiguïté syntaxique en n'effectuant pas de rattachements potentiellement ambigus, tel que les rattachements de groupes prépositionnels.

Modèles de langage à portée variable

A l'issue de cette étape d'analyse nous disposons d'un corpus composé de mots regroupés en syntagmes (de tailles variables) et de mots isolés. La grammaire employée comporte quatre-vingt règles de réécriture.

En remplaçant chaque mot isolé par sa classe et chaque syntagme par une étiquette lui correspondant (GVS pour groupe verbal singulier, par exemple) nous obtenons un corpus dont le vocabulaire est de 70 symboles.

Un modèle 3-gram est ensuite appris sur le corpus analysé. Nous avons repris dans la Table 4 la phrase d'exemple évoquée ci-dessus accompagnée de son analyse.

<i>Mots</i>	<i>Catégorie syntaxique</i>	<i>Syntagme</i>	<i>Décodage</i>
Les	DETMP	GNP	Les
cours	NMP		cours
de	PREP	GP	de
la	DETFS		la
bourse	NFS		bourse
continue <i>ou</i> continuent	V3S <i>ou</i> V3P	GVP <i>ou</i> GVS	continuent
de	PREP	GP	de
s'	PREFMS		s'
effondrer	VINF		effondrer

Table 4 – décodage d'un graphe d'homophone avec le modèle syntagme

On remarque dans cet exemple que du fait du regroupement de certains mots sous forme de syntagmes, le phénomène d'accord entre le groupe nominal sujet et le verbe peut maintenant être capté par un modèle 3-gram. En effet, les trois mots "*de la bourse*" ont été regroupés au sein d'un syntagme prépositionnel, rapprochant ainsi le groupe nominal "*les cours*" du verbe "*continuent*".

Parallèlement, les différentes structures possibles d'un syntagme, en terme d'une simple suite d'étiquettes sont mémorisées. A chaque type de syntagme correspond donc un ensemble de suites d'étiquettes représentant les différentes structures attestées de ce syntagme dans le corpus d'apprentissage. Dans les expériences que nous avons menées, 3000 structures différentes de syntagmes ont ainsi été détectées.

Ces structures constituent une nouvelle grammaire régulière, sous-ensemble de la grammaire utilisée pour l'analyse. La grammaire ainsi créée servira à reconnaître les syntagmes lors de la phase de décodage. Cette identification est effectuée de manière déterministe, en favorisant de façon systématique les syntagmes les plus longs.

Enfin, lors de l'évaluation d'une phrase candidate par le module de stack-decoding, la phrase est étiquetée et découpée en syntagmes avant de recevoir une probabilité calculée comme le produit des probabilités des triplets de syntagmes la composant.

Par exemple, la phrase d'exemple de la table 4 sera découpée selon le schéma suivant :

$W=[GNP \ GP \ GVP \ GP]$ et la probabilité de cette séquence sera calculée comme suit :

$$P_{M_3}(W) = P(GNP|\#, \#) \times P(GP|\#, GNP) \times P(GVP|GNP, GP) \times P(GP|GVP, GP)$$

Bien sûr, ces probabilités sont combinées linéairement avec les probabilités issues des modèles n-class et n-gram. Le score attribué à la phrase par le modèle basé sur les syntagmes

reflète la bonne formation syntaxique de la phrase et en particulier le respect ou la violation des phénomènes d'accord.

3.4. Modèle Homophone-Cache

L'idée originale de cette approche est d'apprendre un modèle discriminant spécifique au traitement des mots homophones singulier/pluriel. En effet, quelle que soit la qualité du modèle de langage traitant les accords entre syntagmes à l'intérieur d'une phrase, il y a de nombreux cas où ceux-ci s'avèrent inopérants. Nous pouvons classer ces situations "difficiles" en trois catégories :

- phénomènes syntaxiques non couverts par la grammaire du modèle syntagme (syntagmes prépositionnels ou propositions relatives imbriquées, coordination, etc.)
- cas indécidables syntaxiquement ou réellement ambigus
- erreur du module de reconnaissance sur la partie de la phrase déjà traitée (insertion, substitution, suppression)

La première catégorie de problèmes nécessite une analyse syntaxique complète de la phrase pour résoudre les ambiguïtés. Par exemple, dans la phrase de la Table 5, l'accord entre

<i>Référence</i>	<i>Catégorie</i>	<i>Syntagme</i>	<i>3-syntagme</i>	<i>cache</i>
les	DETMP	NMP	les	les
principaux	AMP		principaux	principaux
accords	NMP		accords	accords
sur	PREP	GP	sur	sur
l'	DETFS		l'	l'
union	NFS		union	union
douanière	AFS		douanière	douanière
et	COCO	COCO	et	et
pour	PREP	GP	pour	pour
les	DETMP		les	les
six	CHIF		six	six
mois	NMP		mois	mois
la	DETFS	NMS	la	la
monnaie	NFS		monnaie	monnaie
commune	AFS		commune	commune
risquent <i>ou</i> risque	V3P <i>ou</i> V3S	VS <i>ou</i> VP	risque	risquent
d'	PREPADE	PREP	d'	d'
être	VINF	VINF	être	être
bousculés <i>ou</i> bousculé	VPPMP <i>ou</i> VPPMS	PPMS	bousculé	bousculés

Table 5 – décodage d'un graphe d'homophone avec le modèle syntagme et le modèle cache

le verbe « risquent », le participe passé « bousculés » et le groupe nominal « les principaux accords » n'est pas capté par notre modèle syntagme. En effet, ce modèle réalise l'accord entre le verbe et le groupe nominal « la monnaie unique ».

Dans la deuxième catégorie de problèmes, nous rangeons les ambiguïtés réelles, comme :

L'adoption des mesures d'incitations ...

ou encore les ambiguïtés nécessitant des informations lexicales ou sémantiques pour être levées.

Par exemple :

*Le président Boris Eltsine dans un message de vœux **diffusé** à la télévision russe*

L'accord entre "*diffusé*" et "*message*" ne peut être pris en compte par un modèle syntaxique.

Le troisième cas est spécifique au traitement de phrases issues d'un système de reconnaissance de la parole. En effet, la mauvaise reconnaissance d'un mot ("*et*" à la place de "*est*", par exemple) peut rendre difficile tout traitement syntaxique sur la phrase à reconnaître.

En l'absence d'une analyse linguistique complète des phrases à traiter, ces trois catégories de problèmes illustrent le besoin d'un modèle robuste, capable de prendre une décision sur l'ensemble de la phrase déjà décodé sans contraintes contextuelles fortes.

Le modèle proposé consiste à stocker pour chaque flexion singulier/pluriel homophone des lemmes de notre lexique, ses contextes gauches vu dans le corpus d'apprentissage. Ces contextes sont codés sous la forme d'une mémoire cache (Kuhn 1990) représentant un historique limité à 10 mots.

Le cache utilisé est un vecteur dont chaque composante représente une catégorie syntaxique. Le jeu de classes comprenant 105 catégories, chaque vecteur aura ainsi 105 composantes. L'apprentissage d'un tel modèle consiste à parcourir le corpus d'apprentissage en mettant à jour pour chaque occurrence d'une forme la mémoire cache qui lui est associée.

À la fin de l'apprentissage, nous disposons pour chaque homophone de deux vecteurs correspondant à l'historique du mot, un pour la forme singulier de l'homophone et l'autre pour la forme pluriel.

Durant le décodage, lorsque deux flexions homophones singulier/pluriel d'un même lemme sont en compétition, deux distances sont calculées : la distance entre le cache courant et le cache associé à la forme singulier du lemme ainsi que la distance entre le cache courant et le cache de la forme pluriel. La distance utilisée est une distance symétrique, variante de la distance de Kullback-Leibler (Bigi 1998).

Lorsque la différence de ces deux distances est supérieure à un seuil appris sur un corpus de développement, le système choisit la forme correspondant à la distance la plus faible.

La Table 5 montre que le modèle cache permet d'attribuer la bonne catégorie syntaxique au verbe « risquent » ainsi qu'au participe passé « bousculé » en corrigeant les erreurs introduites par le modèle syntagme.

Dans la Table 6, nous présentons un exemple de phrase correctement traitée par le modèle Homophone-Cache. Le modèle a calculé une distance entre, d'une part un vecteur construit à partir des catégories syntaxiques des 9 mots précédant le mot "*diffusé*" dans la phrase (vecteur $A(W)$ de la figure 1) ; et d'autre part les vecteurs caches stockés dans le modèle et correspondant aux flexions *diffusé* (fig1 vecteur C_S) et *diffusés* (fig1 vecteur C_P).

1	2	3	4	5	6	7	8	9	10
Le	président	Boris	Elsine	dans	un	message	de	vœux	diffusé
DETMS	NMS	XPRE	XFAM	PREP	DETMS	NMS	PREP	NMP	diffusés
									VPPMS
									VPPMP

Table 6 – exemple d'utilisation du modèle Homophone-Cache

$A(W)$: vecteur courant de la mémoire cache sur les 9 premiers mots de la phrase

1	2	3	4	102	103	104	105
1	0	1	2	0	0	2	1	2	0

$C_S(W)$: vecteur stocké dans le modèle cache pour la flexion : *diffusé*

1	2	3	4	102	103	104	105
0.8	1.2	2.2	1.6	0.6	1.0	2.8	0.6	0.6	0.2

$C_P(W)$: vecteur stocké dans le modèle cache pour la flexion : *diffusés*

1	2	3	4	102	103	104	105
0.2	0.8	1.2	0.6	2.2	1.8	1.0	0.8	0.1	0.1

Figure 1 – exemples de vecteurs cache utilisés dans le modèle

Dans cet exemple c'est la forme singulier qui est choisie car la différence des distances calculées entre le vecteur A et les vecteurs C_S et C_P est supérieur à un seuil s :

$$|dist_{kl}(A(W), C_S(W)) - dist_{kl}(A(W), C_P(W))| \geq s$$

Le seuil s a été estimé sur un corpus de développement.

4. Évaluation des modèles

4.1. Protocole de test

Nous avons testé nos modèles de langage sur un corpus de textes journalistiques issus du journal *Le Monde Diplomatique* composé de 80K mots. Pour chaque phrase de ce corpus nous avons généré un graphe de mots correspondant aux ambiguïtés d'homophones singulier/pluriel. Enfin, après la phase de décodage, nous avons mesuré le taux de décisions correctes pour ces homophones en comparant la phrase reconnue à la référence initiale.

4.2. Résultats

Les tableaux 7 et 8 montrent les résultats obtenus sur la désambiguïsation de 17.4K homophones singulier/pluriel de notre corpus de test. Nous donnons les résultats en fonction des modèles de langage utilisés : M1=3-gram sur les mots ; M2=3-classe sur les 103 classes syntaxiques ; M3=3-gram sur les syntagmes obtenus après l'analyse grammaticale partielle ; M4=Modèle Homophone Cache.

Modèle	M1	M2	M3	M4
%correct	90.95	95.36	89.02	84.59

Table 7 – résultat sur chacun des modèles

Modèle	M1+M2	M1+M3	M2+M3	M1+M2+M3	M1+M2+M3+M4
%correct	96.89	96.14	96.22	96.98	97.36

Table 8 – résultat sur les mélanges de modèles

4.3. Discussion

Ces résultats montrent que la combinaison de l'ensemble des modèles améliore significativement les résultats. Cela signifie que les dépendances captés par chacun des modèles se complètent.

Il est normal que les performances des modèles M3 et M4 employés seuls soient moins bonnes que les 3-gram ou les 3-class classiques. En effet, le modèle syntagme utilise uniquement 70 classes, il donc moins précis que le autres modèles n-gram.

Le modèle Homophone-Cache, du fait de sa permissivité au niveau des contraintes contextuelles, génère beaucoup de bruit. Il est donc indispensable de le coupler à d'autres modèles pour ne l'utiliser que lorsque ces derniers sont incapables de prendre une décision.

Ainsi, la combinaison de M3 et M4 avec M1 et M2 permet de traiter des problèmes d'accord que les seuls modèles à historique fixe ne peuvent régler.

Cependant la manière de combiner les modèles doit encore être étudiée en détail. En effet, la combinaison linéaire utilisée dans cette étude à le mérite de pouvoir être implémentée simplement en estimant des coefficients pour chaque modèle sur un corpus de développement. Néanmoins, il semble naturel de faire dépendre ces coefficients, non seulement du modèle, mais aussi du type d'homophone traité et du contexte d'apparition de celui-ci. Nous travaillons dans ce sens actuellement pour proposer une stratégie de combinaison des différents modèles basée sur une mesure de la qualité des scores fournis par chacun d'eux face à un contexte particulier.

5. Conclusion

L'objectif de cette étude concerne le traitement de contraintes d'accord entre syntagmes qui ne peuvent être captées par des modèles à portée locales de type 3-gram. Les deux modèles proposés, le modèle à base de syntagme et le modèle Homophone-Cache, permettent de résoudre certaines de ces contraintes d'accord à travers deux aspects différents : le modèle à base de syntagme permet d'introduire des contraintes syntaxiques alors que le modèle Homophone-Cache a pour effet de discriminer les homophones singulier/pluriel de manière robuste, en étant peu sensible à la mauvaise reconnaissance d'un mot au sein de la phrase.

Des tests sont actuellement menés pour confirmer l'intérêt de cette méthode dans le module linguistique d'un système de SRAP à travers deux aspects : l'intégration directe des différents modèles dans le module linguistique de décodage et l'utilisation du modèle Homophone-Cache pour détecter d'éventuelles incohérences de décodage en permettant d'aller chercher de nouvelles hypothèses dans le graphe de décodage.

Références

BÉCHET F., SPRIET T., EL-BÈZE M. Automatic assignment of part-of-speech to out-of-vocabulary words for text-to-speech processing; Eurospeech, Rhodes Grèce, 1997.

BELLEGGARDA J. (1998) *Multi-span Statistical language Modelling for Large Vocabulary Speech Recognition. ICSLP 1998.*

BIGI B., DE MORI R., EL-BÈZE M., SPRIET T. (1998) *Detecting topic shifts using a cache memory, ICSLP'1998*

DELIGNE S., BIMBOT F.(1995). *Language modeling by variable length sequences: theoretical formulation and evaluation of multigrams. ICASSP 95.*

DELIGNE S., SAGISAKA Y. (1998), *Learning a syntagmatic and paradigmatic structure from language data with a bi-multigram model Coling-ACL'98, Montreal, pp 300-306*

GILLET J., WAYNE W. (1998) *A Language Model Combining Trigrams and Stochastic Context-free Grammars. ICSLP 1998.*

KUHN R., DE MORI R (1990) *A Cache Based Natural Language Model for speech Recognition IEEE Trans. Pattern anal. Machine Intell., PAMI-12(6):570-582.*

SPRIET T., EL-BÈZE M. (1995) *Étiquetage probabiliste et contraintes syntaxiques. TALN 95.*

ZITOUNI I, SMAÏLI K., Haton J P., Deligne S., Bimbot F.(1998) *A Comparative Study between Polyclass and Multiclass Language Models. ICSLP 1998.*