

TRAITEMENT DE LA PAROLE

R. Boite, H. Boulard, T. Dutoit, J. Hancq et H. Leich

Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse
1999 — ISBN 2-88074-388-5 — 488 p.

Compte-rendu

La parole articulée — de même que la maîtrise d'un langage ne se résumant pas à une manipulation de symboles en nombre limité — est une des spécificités de la cognition humaine. Aussi l'étude de la parole et du langage parlé a-t-elle toujours été considérée comme un domaine d'investigation privilégié pour les Sciences Cognitives. Il semblerait donc naturel que le Traitement Automatique de la Parole (TAP par la suite) constitue une problématique pluridisciplinaire à l'interface des Sciences du Langage et de l'Intelligence Artificielle. Ce fut ainsi le cas au cours des années 70 et 80, lorsque reconnaissance et synthèse de parole faisaient appel aux compétences d'acousticiens, phonéticiens ou linguistes pour la mise en œuvre de systèmes analytiques à base de règles. Cette modélisation anthropocentrée n'a malheureusement pas tenu ses promesses. Bien au contraire, ce sont des approches "aveugles" basées sur un apprentissage automatique sur de grands corpus (classification bayésienne, modèles de Markov, réseaux de neurones artificiels) qui ont permis l'apparition de systèmes opérationnels au cours de la dernière décennie. Ces succès réels — mais dont on ne doit pas négliger la portée encore limitée — doivent ainsi beaucoup à l'adaptation de techniques issues des Sciences de l'Ingénieur⁹. Aussi ne s'étonnera-t-on pas que cet ouvrage, qui propose un état de l'art très complet des techniques utilisées en TAP, s'adresse explicitement à l'ingénieur.

Le lecteur cognicien doit donc être averti : il ne s'agit pas d'un ouvrage d'introduction approfondie au TAP, comme l'est par exemple le *Natural Language Understanding* de Allen (Allen, 1998) pour le Traitement Automatique du Langage Naturel (TALN par la suite). Il présente au contraire un exposé technique¹⁰ destiné en premier lieu à des personnes disposant — entre autres — de solides connaissances en Traitement du Signal. Nous tenterons donc de faire un compte-rendu à orientation plutôt cognitive de ce livre, tout en gardant à l'esprit les objectifs qui lui ont été assignés.

Le chapitre introductif renseigne le lecteur non averti sur les intentions des auteurs : c'est un parti technique qui est directement adopté. Si on trouve un rappel bienvenu des différents niveaux impliqués dans le traitement de la parole, on regrettera ainsi l'absence de mise en situation de cette problématique. On remarquera également l'attention accordée par cette introduction aux niveaux de traitement supra-lexicaux (syntaxe, sémantique, pragmatique). Nous reviendrons ultérieurement sur cette question.

⁹ Classification statistique de formes, par exemple.

¹⁰ Le dernier chapitre consacré à l'implantation matérielle des algorithmes utilisés en TAP est de ce point de vue emblématique des objectifs de l'ouvrage.

Les deux chapitres suivants concernent la modélisation du signal vocal et sa représentation dans le domaine fréquentiel (spectre du signal). Ils proposent un exposé complet et très précis des techniques correspondantes. Celui-ci nécessite cependant une lecture attentive "crayon à la main". La bibliographie présentée en fin de chapitres se limite à quelques ouvrages, ce qui traduit le caractère assez stabilisé de l'état de l'art en la matière : ne sont généralement présentés que des ouvrages de synthèse bien choisis, qui permettent d'approfondir l'étude de certaines techniques. On regrettera en revanche la présence de coquilles souvent systématiques¹¹.

Le contenu de ces chapitres requiert clairement des connaissances solides en Traitement du Signal et ne sera donc pas accessible aux non spécialistes. Par exemple, le calcul de la transformée en Z d'un signal est supposé connu¹². Il en va de même pour le quatrième chapitre, consacré au codage du signal de parole. Il intéressera avant tout l'ingénieur en télécommunications. De mon point de vue, il est cependant relativement possible à un lecteur non expert de reprendre la lecture de l'ouvrage au chapitre 5, consacré à la Reconnaissance Automatique de la Parole (RAP par la suite).

Ce chapitre présente là encore un exposé très complet (150 pages !) des techniques utilisées en RAP. Sa lecture est cependant assez déroutante. Plutôt que de présenter directement — dans une approche certes très pragmatique — les algorithmes de bases utilisés en reconnaissance de parole¹³, les auteurs choisissent au contraire de replacer la RAP dans la problématique plus générale de la classification de formes. Il s'agit là d'une mise en perspective intéressante : rappeler, par exemple, que les algorithmes d'entraînement de Baum-Welch ou de Viterbi ne sont que des variantes de l'algorithme EM permet de replacer la RAP dans ses fondements théoriques. Cependant, je ne suis pas certain que cette présentation théorique et modélisante facilite une première approche de la reconnaissance de parole¹⁴. Question d'objectifs, une fois encore... Au final, ce chapitre dispose d'une bibliographie là encore très

¹¹ Je pense en particulier aux références, dans le chapitre 3, aux équations du chapitre précédent : celles-ci sont quasiment toutes décalées...

¹² Rappelons, de ce point de vue, que cet ouvrage fait partie d'une collection comportant en particulier des ouvrages sur le Traitement du Signal.

¹³ À l'image par exemple du "petit" *Tutorial on Hidden Markov Models* de Rabiner (Rabiner, 1989).

¹⁴ Je passe sous silence dans ce compte-rendu la partie consacrée à l'apport des réseaux de neurones artificiels en RAP : celle-ci est très bien introduite, tant du point de vue technique que dans l'exposé critique des apports de cette modélisation.

utile. On regrettera plus, dans cette revue défendant l'expression scientifique en langue française, la présence de figures rédigées en anglais et non traduites...

Le chapitre 6 constitue une introduction très rapide au traitement du langage naturel. Il est regrettable que cette synthèse, au demeurant bien faite, se limite à un " kit de survie " en TALN : s'il n'y avait un bref exposé consacré aux grammaires d'unification¹⁵, cette introduction se limiterait en effet à un survol des grammaires régulières et hors-contextes telles que les découvre un étudiant de second cycle en informatique ou en sciences du langage... Les auteurs insistent suffisamment à plusieurs reprises — et à juste titre ! — sur l'insuffisance des modélisations actuelles du langage parlé et sur l'importance de cette question pour l'avenir du TAP, pour qu'on s'attende à en trouver une description beaucoup plus profonde.

Ainsi, aucune mention n'est faite des recherches actuelles sur des modèles de langages qui seraient plus puissants que les N-grams tout en garantissant toujours une interface efficace avec la partie classification statistique de forme de la RAP. De la même manière, la compréhension automatique et le dialogue oral homme-machine de la parole ne sont jamais évoqués dans cet ouvrage.

Il est clair que la question du traitement du langage parlé sort des objectifs assignés à cet ouvrage. Néanmoins, ce genre de restriction ne peut que conforter une vision de la communication parlée limitée à sa dimension infralinguistique. Il existe pourtant une spécificité du langage parlé (Blanche-Benveniste *et al.*, 1990), et de sa modélisation dépend certainement pour une bonne part les progrès futurs des technologies orales.

Notons enfin que c'est à ce niveau que se retrouve désormais déplacé le débat entre modélisation à base de connaissance et modélisation purement statistique de la parole : si les modèles N-grams ont permis d'obtenir des résultats appréciables en terme de robustesse, leurs limitations sont cependant de plus en plus manifestes. Ainsi, il n'est pas interdit de penser que la recherche d'applications plus complexes pour les technologies orales¹⁶ puisse se traduire par le retour d'approches plus motivées linguistiquement (Antoine et Genthial, 1999)¹⁷. Cette évolution possible du Traitement de la Parole, éludée dans cet ouvrage, concerne donc au premier chef les Sciences Cognitives.

Ce manifeste en faveur d'une prise en compte de la dimension langagière de la parole ne doit pas faire oublier le chapitre 7 — conséquent (97 pages) tout

comme sa bibliographie — consacré à la synthèse de la parole à partir d'un texte¹⁸. Issu, comme le chapitre précédent, de la traduction d'un ouvrage anglophone (Dutoit, 1997), il tranche par sa forme avec le reste du livre : tout en ne délaissant pas l'aspect modélisation cher à l'ingénieur, il propose un panorama beaucoup plus synthétique de cette problématique. Le lecteur novice peut ainsi avoir une vision assez globale des traitements mis en œuvre, des approches utilisées pour les réaliser et de leurs enjeux.

À ce sujet, on peut s'interroger sur la diversité des approches utilisées à l'heure actuelle en synthèse : alors que les systèmes de RAP actuels sont fondés sur le même substrat théorique (classification bayésienne et modélisation markovienne du langage), on distingue ici de nombreuses approches entre synthèse par règle, ou synthèses par concaténation diverses (LPC, modélisation harmonique/stochastique, temporelle TD-PSOLA ou MBROLA...). D'où la question : cette diversité est-elle due à un manque de maturité des recherches dans le domaine¹⁹ ou au contraire à la plus grande difficulté (mais de quel point de vue ?...) de la tâche. Il me plaît de penser que la seconde réponse est la bonne : la perception humaine doit certainement être impitoyable vis à vis du signal de parole synthétisé, d'où les difficultés rencontrées par une synthèse de parole ne faisant appel qu'aux Sciences de l'Ingénieur.

À l'opposé, on reste étonné par la grande tolérance aux erreurs que manifestent les utilisateurs des systèmes de RAP ou de dialogue oral homme-machine. Si cette tolérance demande à être mieux étudiée²⁰, elle explique certainement pour partie les réussites rencontrées par des systèmes basés sur une modélisation assez fruste du langage parlé. Le développement de technologies orales grand public passant par une tolérance sensiblement moindre, l'apport d'approches linguistiquement plus réalistes pourrait alors se faire plus pressant. Certains travaux (Chelba et Jelinek, 2000, pour ne prendre qu'un exemple emblématique) tendent à montrer que cette question n'est pas dénuée de fondements, même dans une optique clairement d'ingénierie.

Ainsi, quinze ans après la " défaite analytique ", les Sciences Cognitives pourraient encore avoir leur place dans la recherche en TAP...

Ce type d'interrogation épistémologique est totalement absent de cet ouvrage qui se concentre sur l'exposé des techniques actuellement mises en œuvre en TAP. De ce point de vue, il décevra le cognicien à la recherche d'une introduction synthétique et critique sur ce domaine de recherche. Il ne faudrait cependant pas en oublier les

¹⁵ Mais pas, par exemple, à des formalismes dérivés récents tels que les grammaires d'arbres adjoints (grammaires TAG).

¹⁶ Par exemple, comprendre la parole spontanée et non plus seulement la reconnaître, et ceci dans des contextes beaucoup moins finalisés que ceux étudiés à l'heure actuelle (ATIS par exemple) en CHM orale.

¹⁷ Tout en étant compatibles avec des méthodes issues des Sciences de l'Ingénieur. On notera à ce sujet que la plupart des formalismes actuellement utilisés en TALN (grammaires TAG, grammaires de liens ou de dépendances) a déjà donné lieu à des mises en œuvre probabilisées et que la question de l'inférence automatique de telles grammaires reste un sujet " chaud " du domaine.

¹⁸ On notera là encore que le dialogue oral homme-machine, qui peut donner lieu à une synthèse à partir de concepts et non à partir du texte, n'est une fois encore qu'évoqué (p. 430).

¹⁹ Soit qu'une approche réellement satisfaisante n'ait pas encore émergé, ou, ce que tendrait plutôt à montrer le texte, qu'on ne disposait pas encore jusqu'à un passé récent de bases de données étiquetées suffisamment importantes pour bénéficier de toute la puissance d'approches aveugles de type " force brute " (apprentissage statistique ou neuronal).

²⁰ Quid de l'utilisateur réel, dans une situation non artificielle, et disposant d'un choix entre interaction avec la machine ou un opérateur humain ?

qualités de ce livre, qui nous donne en près de 500 pages un panorama très complet des technologies orales. Présenté comme un ouvrage destiné à l'ingénieur, il ne fait nul doute qu'il répondra également aux attentes des chercheurs ou étudiants de troisième cycle à la recherche d'un exposé détaillé sur le Traitement de la Parole. Jusqu'ici, le chercheur francophone ne disposait réellement que du *Calliope* (Calliope, 1989) comme ouvrage de référence. Cette dernière parution le complétera très utilement et y apportera un regard renouvelé par dix années de recherche.

À l'étudiant à la recherche de références introductives solides, je conseillerais cependant d'aborder tout d'abord le *Calliope*, pour ensuite approfondir ses connaissances avec cet ouvrage faisant une part plus grande aux fondements théoriques du TAP. La lecture devrait alors en être plus digeste !

Références bibliographiques

[Antoine et Genthial, 1999] Antoine J.-Y., Genthial D. (1999). Méthodes hybrides issues du TALN et du TAL Parlé : état des lieux et perspectives, actes *TALN'99*, atelier thématique "Méthodes hybrides TALN / TALP pour le traitement robuste de la langue, Cargèse, France, 1-17.

[Allen, 1998] Allen J. (1998). *Natural Language Understanding*. Benjamins Cummings : New-York. 2nd édition.

[Blanche-Benveniste et al., 1990] Blanche-Benveniste C., Bilger M., Rouget C., Van den Eynde K. (1990). *Le français parlé : études syntaxiques*. CNRS Éditions : Paris.

[Calliope, 1989] Tubach J. P. (ed.). 1989. *La parole et son traitement automatique*. Masson : Paris.

[Chelba et Jelinek, 2000] Chelba C., Jelinek F. (2000). Structured language modeling. *Computer Speech and Language*. 14(4), October. 283-332.

[Dutoit, 1997] Dutoit T. (1997). *An introduction to Text-to-Speech Synthesis*. Kluwer : Dordrecht.

[Rabiner, 1989] Rabiner L. R. (1989). A Tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*. 77 (2). 257-285.

L'auteur de la revue critique



Jean-Yves Antoine est maître de conférences en informatique à l'Université de Bretagne Sud. Après un doctorat sur la compréhension de parole préparé à l'Institut de la Communication Parlée (Grenoble), il a mené des études post-doctorales sur le même thème au CLIPS-IMAG (Grenoble) avant de rejoindre le

laboratoire VALORIA de l'Université de Bretagne Sud, à Vannes. Il dirige actuellement le groupe de recherche en ingénierie linguistique de ce laboratoire, où il conduit des travaux sur la compréhension de la parole en dialogue homme-machine, l'aide linguistique aux handicapés, l'évaluation des systèmes de dialogue ainsi que des recherches en linguistique de corpus. Il anime également un groupe de recherche du PRC-I3 sur la compréhension robuste de la langue. Il est enfin rédacteur en chef de la revue en sciences cognitives *In Cognito*.

La réaction des auteurs

Comme le constate la revue critique ci-dessus, nous n'avons pas précisément écrit ce livre pour un lecteur cognicien. Il s'agit en effet d'un ouvrage dans lequel nous avons voulu brosser l'état de l'art en matière de traitement de parole, de la façon la plus complète possible et en un seul volume (l'une et l'autre de ces contraintes étant bien évidemment antagonistes). Les techniques décrites dans cet ouvrage sont donc celles qui sont réellement utilisées aujourd'hui dans des produits commerciaux ou en voie de commercialisation. Il est clair, pour nous qui sommes confrontés tous les jours aux demandes des industriels de la parole, que l'approche analytique a de moins en moins d'impact sur les technologies de codage, synthèse et reconnaissance de la parole, au profit des approches statistiques, basées sur l'entraînement de machines génériques à partir de grandes bases de données de texte et de parole.

C'est un fait. Nous n'avons fait qu'en rendre compte. Comme signalé au chapitre sur la synthèse : "après tout, les avions ne battent pas des ailes"...

Les auteurs