
SPOKEN LANGUAGE PROCESSING

A GUIDE TO THEORY, ALGORITHM AND SYSTEM DEVELOPMENT

X. HUANG, A. ACERO, H.-W. HON

2001 — Prentice Hall, Upper Saddle River, NJ — ISBN 0-13-022616-5 — 980 pp. — 89,99 \$

Fin 1999 paraissait aux Presses Polytechniques et Universitaires Romandes un ouvrage francophone intitulé « Traitement de la Parole » (Boite *et al.*, 1999). Son ambition était de rendre compte des théories et techniques mises en jeu dans les applications actuelles relevant de l'ingénierie de la parole. Ouvrage de référence très complet, ce livre se limitait cependant au traitement de la parole *stricto sensu*, et n'évoquait que brièvement le traitement du langage parlé et ses applications. Le lecteur intéressé par ces niveaux de traitement linguistique devait alors s'en remettre à des ouvrages spécifiques tels que celui d'Allen (1995) par exemple.

Comme son titre le suggère, le « *Spoken Language Processing* » de Huang, Acero et Hon s'intéresse au contraire à cette dimension linguistique, envisagée dans le cadre du dialogue homme-machine oral (ou multimodal). Son propos est cependant plus ambitieux. Il présente en effet, dans un exposé de près de 1000 pages ne souffrant d'aucune lacune, l'ensemble des niveaux de traitement de la parole.

Après une brève introduction, les auteurs s'attardent tout d'abord sur les connaissances théoriques de base nécessaires à la compréhension des chapitres suivants. Relevant à la fois des sciences du langage (phonétique, syntaxe, sémantique), de la statistique (théorie de l'information) et de l'Intelligence Artificielle (reconnaissance des formes), cette introduction bien conduite et jamais sommaire (200 pages) rappelle le caractère pluridisciplinaire du traitement du langage parlé. On se félicitera au passage que les auteurs aient rattaché dans cette introduction la théorie de l'estimation aux tests de significativité statistique (test de la normale, du Chi-deux par exemple). Reconnue en psychologie expérimentale, l'importance de ces tests est en effet souvent sous-estimée par les chercheurs en parole travaillant sur des approches statistiques.

Après cette présentation des fondamentaux en traitement de la parole, l'ouvrage va s'intéresser successivement aux grands champs d'application du domaine. Logiquement, le chapitre suivant est consacré au traitement du signal de parole. Bien structuré (traitement du signal proprement dit, représentation puis codage du signal de parole), ce chapitre est particulièrement complet. À titre d'exemple, il s'attarde dans un premier temps sur les transformées de Fourier et en Z continues, là où l'ouvrage de Boite *et al.* (1999) se contente d'aborder directement l'analyse de Fourier discrète à court-terme. Comme dans cet autre ouvrage, les démonstrations ne sont le plus souvent qu'esquissées, et aucun exercice n'est proposé au lecteur. La lecture de ce chapitre requiert donc une forte attention — stylo et papier en main — et il est à craindre qu'il ne soit que difficilement accessible à des lecteurs non scientifiques. Il est cependant clair qu'un traité complet en traitement du signal n'aurait pas sa place ici,

et que l'ouvrage présente déjà suffisamment d'éléments pour guider le non-spécialiste vers des lectures plus approfondies.

De ce point de vue, les synthèses bibliographiques et historiques proposées à la fin de tous les chapitres sont particulièrement bienvenues. Les références bibliographiques retenues, qui se partagent le plus souvent entre ouvrages de synthèse à visées pédagogiques et articles phares ayant marqué l'évolution du domaine, paraissent ainsi toujours utiles et pertinentes.

Le chapitre suivant, consacré à la reconnaissance de la parole, bénéficie de la grande expérience en la matière des auteurs. Tous ont en effet rejoint le *Speech Technology Group* de Microsoft Research après des recherches menées à Carnegie Mellon (CMU). C'est là l'occasion de rappeler l'importance et la pertinence des travaux de Microsoft Research dans le domaine des technologies langagières, mais aussi de se féliciter que les auteurs aient su résister à la tentation de centrer leur ouvrage sur les seules recherches et applications commercialisées par ce groupe. Avec plus de 300 pages, ce chapitre est le plus conséquent de l'ouvrage. Il fournit une présentation détaillée des techniques utilisées en reconnaissance, que ce soit pour la modélisation acoustique ou la modélisation du langage. On regrettera que l'utilisation de réseaux de neurones pour la modélisation acoustique n'est que brièvement évoquée ici, à l'opposé de l'ouvrage de Boite *et al.* (1999), dans lequel la place significative accordée aux approches connexionnistes s'explique certes par la présence parmi les auteurs d'Hervé Bourlard, promoteur de ces techniques en reconnaissance de parole. Les résultats significatifs obtenus par ses approches atypiques mériteraient néanmoins un développement plus marqué dans le présent ouvrage.

Ce chapitre consacre également une partie bien identifiée à l'étude des principaux algorithmes de recherche. Cette structuration est bienvenue, tant il est vrai que la reconnaissance de la parole est autant, voire plus, un problème de recherche d'une solution optimale dans un large espace de possibilités qu'un « simple » problème de reconnaissance des formes.

Le quatrième chapitre de l'ouvrage est consacré à la synthèse de parole à partir du texte. Bien que n'étant pas spécialiste du domaine, cette partie m'est apparue une fois encore remarquablement étoffée.

La structuration de l'exposé apparaît une fois encore très judicieuse et pédagogique. En particulier, la prosodie se voit accorder un sous-chapitre bien identifié. Mise en avant pertinente lorsque l'on sait que la génération de la prosodie constitue actuellement une des pierres d'achoppement du domaine. Les problèmes d'intelligibilité dus au caractère métallique des premières voies synthétisées sont désormais relativement bien

maîtrisés (d'Alessandro et Tzoukermann, 2001). C'est donc bien de la naturalité (et de l'expressivité) de la prosodie générée que semblent dépendre les réussites futures de la synthèse de parole.

Ce chapitre ne me laisse qu'un seul regret : l'absence de référence faite à une étape préalable à la synthèse de parole, à savoir la génération du message à synthétiser. La génération automatique de texte ne constitue pas une problématique propre au traitement du langage parlé. Il me semble cependant regrettable de ne pas l'évoquer ici. Les progrès de la reconnaissance de parole ont montré que certaines limitations des systèmes de dialogue oral provenaient de niveaux de traitement supérieurs (compréhension de parole, gestion du dialogue). De même, il est à craindre que les avancées de la synthèse de parole mettront à jour les insuffisances des modules *ad hoc* de génération de réponse actuellement utilisés en communication homme-machine.

Néanmoins, il faut rendre justice aux auteurs : la génération de réponse est — rapidement : 6 pages — évoquée dans le dernier chapitre de l'ouvrage consacré au traitement du langage parlé (*Spoken language systems*). Comme nous l'avons dit, un des principaux intérêts de cet ouvrage est qu'il ne se limite pas aux niveaux infra-linguistiques de traitement de la parole. Motivé par la mise en œuvre des systèmes de dialogue oral homme-machine¹³ — un sous-chapitre est d'ailleurs spécifiquement consacré aux applications typiques de ces systèmes — il étudie ainsi la compréhension de parole, la génération d'une réponse orale et beaucoup plus longuement le contrôle du dialogue entre le système et l'utilisateur. On peut regretter la portion congrue accordée à la compréhension de parole. Cependant, cette situation est simplement révélatrice de l'absence d'intérêt actuel de la communauté scientifique pour cette problématique. La compréhension est en effet le plus souvent perçue comme une simple étape d'interface entre les deux étapes cruciales que sont la reconnaissance de parole (robustesse en entrée) et le contrôle du dialogue (pertinence de l'interaction avec l'utilisateur). À mesure que les systèmes de dialogue oral s'intéresseront à des domaines applicatifs plus complexes (moins finalisés), il est cependant à prévoir que l'importance de ce niveau de traitement sera mieux reconnue.

Au final, cet ouvrage d'ambition encyclopédique se positionne d'emblée comme une référence en matière de technologies orales. Avec un grand souci pédagogique, il propose aux lecteurs une revue de l'état de l'art qui ne présente aucune lacune majeure et reste toujours au fait des dernières évolutions du domaine.

Trop précis pour constituer un simple ouvrage d'introduction aux technologies orales, il ne peut être — par manque de place — assez détaillé pour permettre une compréhension complète des techniques exposées. Il lui faudrait pour cela plus d'exemples illustratifs, de démonstrations détaillées et d'exercices. Au contraire, il répond parfaitement à la définition de l'ouvrage de référence, c'est-à-dire celui que l'on ne lit pas du début à

la fin, mais dans lequel on va chercher ou retrouver une information précise avec l'assurance de la trouver.

En fournissant un survol déjà précis de chaque problématique, survol qui pourra être utilement complété par les références des mises en perspectives de chaque fin de chapitre, il constitue un précieux ouvrage d'entrée qu'on ne peut que recommander à tout chercheur désireux de découvrir ces thématiques, et ce, qu'il s'agisse du doctorant débutant ou du cogniticien confirmé soucieux d'une ouverture pluridisciplinaire. Nul doute donc qu'il accompagnera les propos de nombreux enseignants-chercheurs du domaine.

Références bibliographiques

[d'Alessandro et Tzoukermann, 2001] d'Alessandro C., Tzoukermann E. (2001). Synthèse de la parole à partir du texte. *Traitement Automatique des Langues, TAL*. Hermès : Paris, France. 42(1).

[Allen, 1995] Allen J. (1995). *Natural Language Understanding*. The Benjamin/Cummings Publ. Company : Menlo Park, CA. 2nd édition.

[Boite et al., 1999] Boite R., Bourlard H., Dutoit T., Hancq J., Leich H. (1999). *Traitement de la parole*. Collection *Electricité*. Presses Polytechniques Universitaires Romandes : Lausanne, Suisse.

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-13 sur la compréhension robuste de la langue. Il est enfin rédacteur en chef de la revue en sciences cognitives *In Cognito — Cahiers Romains de Sciences Cognitives*.

¹³ Le dialogue oral homme-homme médiatisé par l'ordinateur, comme dans la problématique émergente de la traduction parole-parole, n'est pas étudié dans cet ouvrage.