PHOLIA 1 (1984), pp. 189-197.

POUR L'APPLICATION AU SWAHILI DES TECHNIQUES DE TRAITEMENT AUTOMATIQUE DE LA PAROLE'

Jean-Marie HOMBERT, François NSUKA NKUTSI
et Gilbert PUECH

1 Le Traitement Automatique de la Parole (TAP) ouvre des perspectives tout à fait nouvelles dans la mesure où les machines entendantes et parlantes sont appelées à devenir des prolongements de l'home comme les outils le sont de ses mains. Grâce à lui le dialogue en langue naturelle s'engage avec la machine : celle-ci joue un rôle de conseil pendant le dialogue et exécute ensuite les directives reçues oralement. Les investissements massifs des grands de l'Electronique et de l'Informatique, l'organisation de filières de recherches sous l'égide des Pouvoirs publics (projet ARPA aux Etats-Unis, GRECO sur la parole en France) montrent bien que le TAP est considéré comme un enjeu important pour l'avenir. La collaboration active entre l'Industrie et la Recherche a permis des progrès importants dans la modélisation de la parole mais aussi la mise au point de circuits intégrés pour le traitement de la parole à des prix qui permettent d'envisager des applications accessibles à tous dans un avenir rapproché.

Les pays industrialisés commencent à appliquer le TAP dans des secteurs où la commande vocale remplace avantageusement

---

[1] Communication présentée à la Table-Ronde Internationale du C.N.R.S. "Le Swahili et ses limites : ambiguïté des notions reçues", Sèvres, 20-22 avril 1983.

la commande manuelle ou dans lesquels la parole est le support naturel des opérations effectuées. On peut citer par exemple :

- l'aide à la prise de décision (ex. : dans les procédures d'atterrissage et de décollage),
- la commande vocale de machines, robots, etc.,
- la saisie d'informations orales (ex. : la saisie orale du code postal pour le tri),
- la diffusion de messages actualisés en fonction d'un environnement,
- la consultation en langue naturelle de banques de données sur des réseaux télé-informatiques,
- l'aide à l'apprentissage (orthographe, mathématiques, langues étrangères, etc.),
- l'aide aux handicapés (aveugles, mal entendants, handicapés moteurs).

Des compagnies comme IBM ont des objectifs à plus long terme : machine à écrire automatique ; de même les avancées dans le domaine du TAP conduisent à élargir le champ des recherches en traduction automatique.

Les pays industrialisés ont défini les applications en fonction des besoins qui étaient les leurs. A lire ce qui précède, le TAP peut ne pas apparaître comme une priorité urgente pour les pays du Tiers-Monde. L'enjeu est pourtant d'importance : les "outils" nouveaux développés par la technologie des dix années à venir seront-ils accessibles uniquement par l'intermédiaire des quelques langues "occidentales" de grande diffusion (anglais et, à un moindre degré, Japonais, allemand ou français) ou seront-ils adaptables aux langues nationales ? Prenons l'exemple de la machine à écrire automatique (horizon 199012000) : il est évident qu'un tel outil, s'il n'est pas multilingue, renforcera considérablement la domination des langues étrangères sur une langue nationale pour tout ce qui touche à la communication institutionnelle, avec toutes los conséquences linguistiques et sociologiques que cela entraîne.

C'est dans ce contexte qu'il faut, à notre sens, situer la responsabilité des linguistes, notamment ceux des pays en voie de développement, sur la question du TAP. En se formant à ces nouvelles techniques et en utilisant leur compétence pour adapter les outils aux langues de leur culture, ils feront progresser la recherche fondamentale (qui ne peut se satisfaire des solutions trouvées en fonction de 2 ou 3 langues exclusivement) et joueront tout leur rôle pour aider à tracer de nouvelles voies de développement :

1) En contribuant à définir les applications du TAP utiles à la société dans laquelle ils vivent. Les pays du Tiers-Monde ont généralement des cultures où le rapport entre l'oral et l'écrit est différent de celui qui s'est établi dans les pays hyper-industrialisés. D'un côté l'oralité est une valeur à garder comme une richesse pour la culture, de l'autre l'alphabétisation est une nécessité pour le développement économique. C'est sans doute dans ce contexte qu'il faut repenser les contributions du TAP au développement de pays du Tiers-Monde(ainsi une application qui a donné lieu à la commercialisation d'un jouet pour l'apprentissage de l'orthographe des mots usuels -le Speak and Spell de Texas Instrument- peut être repensée pour devenir une aide efficace à l'alphabétisation, etc).

2) En acquérant les connaissances techniques nécessaires pour concevoir, écrire et tester eux-mêmes les logiciels permettant le TAP dans leur langue.

Les chercheurs qui commencent aujourd'hui à s'intéresser au TAP bénéficient certes des recherches antérieures et des acquis technologiques disponibles. Mais il faut plusieurs années entre le moment où l'on aborde un tel domaine et celui où l'on est vraiment opérationnel et compétent par rapport à l'état de l'art.

Notre équipe de bantouisants a choisi de s'intéresser au swahili et de défricher les problèmes que pose cette langue

pour le traitement automatique de la parole dans notre laboratoire à Lyon. Le choix de cette langue a été dicté par quatre considérations :

1) c'est la langue africaine qui a la plus grande diffusion,

2) c'est une langue bantoue dont la structure, en l'absence de tons, se prête mieux que d'autres à la synthèse et à la reconnaissance automatiques,

3) l'existence de structures universitaires développées dans plusieurs pays swahilophones doit permettre une coopération active avec les linguistes de ces pays,

4) le TAP représente une technologie d'avenir qui peut conforter le statut du swahili comme langue nationale et apporter une aide utile au développement.

Notre démarche implique donc que se mette en place une collaboration avec 1) les spécialistes (européens ou africains) de swahili, en particulier pour ce qui touche à la phonologie, la prosodie, la variation dialectale et la socio-linguistique, 2) des phonéticiens prêts à travailler en laboratoire et sur le terrain pour l'expérimentation des résultats (contrôle de qualité et impact sociologique).

2. Le Traitement Automatique de la Parole se divise en deux branches :

- la Synthèse Automatique de la Parole (SAP),
- la Reconnaissance Automatique de la Parole (RAP).

Dans les deux cas on a une liaison entre un ordinateur (ou un micro-processeur pour les cas simples) et un matériel de traitement du son. L'organe de liaison est un convertisseur :

1) analogique/numérique, s'il s'agit de convertir une onde sonore en une suite de nombres traitables par l'ordinateur;

2) numérique/analogique, s'il s'agit de convertir une suite de nombres en une série d'impulsions électriques formant une onde sonore.

2.1 **Sous** le terme générique de Synthèse on regroupe deux approches complémentaires :

1 Création de parole **ex** nihilo : on transmet des paramètres numériques à un ordinateur qui calcule une suite de nombres (formant une onde numérique) convertie en une onde sonore. Les caractéristiques de la voix du synthétiseur sont fixes et indépendantes du message véhiculé. **Une** forme élaborée de cette approche consiste à partir d'un texte écrit et à le convertir en parole continue.

2. Recréation de parole à partir de l'analyse d'un **message**. Le synthétiseur garde alors les principales caractéristiques de la **voix** émettrice. Cette approche permet **un** stockage très condensé de la parole **(en** divisant le nombre d'informations nécessaire à sa restitution par un facteur de 1 à 100) **et/ou** un **traitement** qui le modifie.

Le passage d'un texte écrit à **une** parole continue implique plusieurs étapes d'analyse :

a) Passage de la transcription orthographique à la transcription phonétique. A **ce** stade-là une analyse morpho-syntaxique peut **se** révéler nécessaire pour lever les ambiguïtés de graphie. Ainsi dans :

"les poules du couvent couvent"

la séquence -ent équivaut d'abord à /ã/ puis à zéro, [kuːv]. **Dans** le **cas** du **swahili,** dont **l'orthographe** s'apparente beaucoup plus à une transcription phonologique que dans des langues comme le **français ou l'anglais, ces** problèmes sont relativement mineurs. En revanche, la détermination de marqueurs prosodiques -nécessaires pour l'obtention d'une parole naturelle à partir d'un texte même ponctué nécessite d'importantes recherches linguistiques sur la structure de l'intonation en swahili.

b) Passage de la transcription phonétique aux paramètres de commande. Plusieurs techniques existent. :

- synthèse Par règles : à chaque phonème correspond un Jeu de paramètres stockés en mémoire. La difficulté est alors d'écrire des règles qui permettent de joindre les phonèmes entre *eux* : on Sait en effet que les transitions formantiques jouent un rôle essentiel dans la perception de la parole continue et notamment pour l'identification du lieu d'articulation des consonnes.
- synthèse Par diphonèmes : on met en mémoire les paramètres de voyelles pures et de voyelles jointes à une consonne. On forme la syllabe en assemblant ces divers éléments :

ex. : #m-mi-i-it-ti-i#
1 2 3 *4 5 6*

Dans le cas du swahili il faut paramétrer 250 diphonèmes environ.

- synthèse par mots : on stocke des jeux de paramètres correspondant à des mots entiers. Le passage de l'écrit à l'oral se réduit alors à l'introduction des marques prosodiques du discours mais naturellement on est limité par un vocabulaire fixe et limité.

c) Production de la parole. Il existe plusieurs techniques :

Le vocoder (voice coder) à filtres comprend un étage d'analyse et un étage de synthèse. A l'analyse le signal, dont on calcule le fondamental, passe par un banc de filtres qui fournit le spectre en fréquence. L'étage de synthèse effectue le processus symétrique.

Synthétiseurs à formants : les synthétiseurs à formants sont constitués de filtres résonants dont la courbe de réponse globale en fréquence reproduit celle du conduit vocal.

Synthétiseurs articulatoires : on modélise la forme du conduit vocal et on calcule l'onde à partir de ses caractéristiques.

Synthétiseurs à codage prédictif : c'est une méthode de simulation de la fonction de transfert du conduit vocal qui consiste à

calculer la valeur de chaque échantillon en fonction de la valeur pondérée (par un jeu de paramètres variables) de N échantillons qui le précèdent.
Synthèse par forme d'onde : on utilise des méthodes de codage numérique pour caractériser les formes d'onde et on peut ainsi les reproduire.

Ces diverses approches sont toutes l'objet de recherche* en laboratoire et donnent lieu à la mise au point de circuits intégrés qui effectuent les calculs très complexes qu'elles nécessitent en temps réel (c'est-à-dire que le temps de calcul d'une fraction d'onde (13 ms par exemple) est inférieur à sa durée, ce qui permet l'obtention d'une parole continuel.

A Lyon nous sommes équipés d'un synthétiseur à codage prédictif (Prosit 2000) utilisant un circuit du CNET. Il faut souligner que le prix de revient de ces appareils, déjà raisonnable, est appelé à diminuer, ce qui permettra de mettre en concurrence plusieurs approches en comparant la qualité de la parole produite et les contraintes d'utilisation.

22 La Reconnaissance Automatique de la Parole doit permettre à l'ordinateur de comprendre un message oral et de poursuivre la tâche qui lui est assignée en fonction du message repu. La difficulté vient de ce que le signal de la parole est continu et variable. d'où des problèmes considérables de segmentation et d'identification des éléments segmentés. On peut contourner cette difficulté en limitant, dans un premier temps, son ambition à des appareils monolocuteurs reconnaissant des mots isolés :

On qualifie de "monolocuteurs" les machines dont l'utilisation passe par deux phases :

1. La phase d'apprentissage par la machine des caractéristiques d'une voix ; cette phase consiste à enregistrer plusieurs fois quelques mots déterminés. Les caractéristiques pertinentes sont analysées et stockées en mémoire. On élimine

ainsi une bonne partie des causes potentielles de variation (différences d'accent dialectal, types de voix - masculin / féminin - etc.). Si un autre locuteur se sert de la même machine il faut recommencer la phase d'apprentissage.

2. La phase de reconnaissance de mots d'un lexique (de 10 à 100 mots par exemple pour une machins simple) ; il est évident qu'en utilisant des mots isolés on évite le problème de la segmentation. La machine sélectionne dans le lexique le mot stocké dont le pattern est le plus en rapport avec le mot oral reçu. On peut également employer des mots connectés (possibilité de plusieurs combinaisons à partir d'un petit nombre de mots élémentaires ; par exemple, les chiffres dans les syntagmes numéraux).

Une machine "multilocuteurs" doit être capable de reconnaître le sens d'un message indépendamment des caractéristiques idiosyncratiques des locuteurs. Cela implique des recherches phonétiques plus poussées et aussi une bonne connaissance des variations attestées dans la communauté linguistique.

Un des critères importants d'évaluation des machines de RAP est le taux d'erreur admis : on considère généralement *que* le taux d'erreur doit être inférieur à 10 % pour qu'on puisse parler de reconnaissance automatique ; ce taux doit être ramené à 1 % pour les applications Industrielles et scientifiques.

On est encore très loin actuellement de pouvoir reconnaître automatiquement la parole continue. Toutefois les industriels mettent au point des circuits de plus en plus performants (avec des prix en diminution rapide), ce qui laisse entrevoir la possibilité de progrès importants dans un avenir proche. Mais là encore des recherches complémentaires sur chaque langue concernée par la RAP sont nécessaires avant que l'on puisse espérer segmenter et catégoriser automatiquement le signal.