

Automates synchronisés pour l'intégration de techniques de reconnaissances de la parole et de compréhension du langage naturel

Patrice Lopez, Jean-Luc Husson

► **To cite this version:**

Patrice Lopez, Jean-Luc Husson. Automates synchronisés pour l'intégration de techniques de reconnaissances de la parole et de compréhension du langage naturel. Atelier TALN Compréhension robuste de la langue: méthodes hybrides entre TALN et TALP, Jul 1999, Cargèse, Corse, France, 10 p. inria-00107533

HAL Id: inria-00107533

<https://hal.inria.fr/inria-00107533>

Submitted on 19 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automates synchronisés pour l'intégration de techniques de reconnaissance de la parole et de compréhension du langage naturel

Patrice Lopez et Jean-Luc Husson

LORIA UMR 7503
BP 239
F-54506 Vandoeuvre-les-Nancy
{lopez, husson}@loria.fr

Résumé

Nous présentons une architecture dont l'objectif est d'aboutir à une intégration forte des différents niveaux de traitement de la langue parlée. Les connaissances statiques sont représentées sous forme d'automates à états finis permettant un partage optimal des sous-structures communes. Ces automates sont utilisés dans la mise en œuvre d'analyses stochastiques et tabulaires afin de prendre en compte le non-déterminisme des différents niveaux de traitement. Des fonctions de synchronisation sont appliquées sur ces automates afin de propager les différentes contraintes entre niveaux. L'architecture obtenue permet d'isoler compétences symboliques et probabilistes, interfaces entre niveaux d'analyse et contrôles. Des expérimentations reposant sur ces principes et ces représentations sont actuellement menées en vue d'intégrer un système de segmentation analytique, un module de reconnaissance phonétique stochastique et un analyseur basé sur les LTAG synchrones.

1. Introduction

Durant des années, l'étude de la langue s'est concentrée sur sa forme écrite, considérée comme norme, soulevant avant tout des questions de *compétence linguistique*, c'est-à-dire s'interrogeant sur les formes linguistiques devant être considérées comme correctes. La langue parlée, bien qu'étant la forme la plus naturelle de la langue, semble avoir longtemps été considérée comme un sous-produit de la forme écrite. Les formalisations automatisables proposées dans ce cadre sont avant tout de nature symbolique, nécessitant d'énormes efforts de modélisation de la part de linguistes afin d'aboutir à des implantations effectives à large couverture langagière, comme le montre par exemple l'implantation du formalisme GPSG (Briscoe *et al.*, 1987).

A l'inverse, les travaux concernant le domaine de la Reconnaissance Automatique de la Parole (RAP) sont soumis à de fortes contraintes d'efficacité relatives à la nécessité pour les systèmes auxquels ils sont destinés de fonctionner en temps réel, et se sont donc davantage focalisés sur les questions de *performance linguistique*, c'est-à-dire sur les formes linguistiques

effectivement observées en conditions réelles. Les solutions incontournables dans ce domaine sont les techniques stochastiques qui permettent d'aboutir à de bonnes performances de reconnaissance au prix d'un processus d'apprentissage sur de très volumineux corpus étiquetés.

La combinaison des technologies de reconnaissance de la parole et de compréhension du langage naturel peut améliorer le potentiel mutuel de ces deux technologies. D'une part, les aspects perceptifs de la parole, par exemple des informations de natures prosodiques (durée et *pitch*), peuvent améliorer sa compréhension en participant à la désambiguïsation des analyses concurrentes, détectant par exemple les frontières de syntagmes, les pauses, etc. De nombreux travaux ont montré les interactions entre syntaxe et phonologie (Selkirk, 1984). D'autre part, en plus d'assigner une description de la structure d'un énoncé et une interprétation, la syntaxe et la sémantique peuvent en retour améliorer les performances d'un système de reconnaissance de la parole en réduisant dynamiquement les choix possibles d'identification de mots.

De nombreux systèmes de dialogue adoptent une architecture séquentielle purement ascendante. Les mécanismes de compréhension se fondent alors souvent sur des formes simples de dialogue, de type informatifs (ATIS¹), en ne cherchant à analyser que les informations susceptibles de faire avancer la tâche (méthodes selectives (Luzzati, 1989)). Or pour des dialogues plus complexes (dialogues de commande ou d'assistance), il est nécessaire de faire intervenir l'ensemble des informations véhiculées dans les énoncés des utilisateurs et de faire appel à des techniques non-déterministes, plus souvent employées pour le traitement de l'écrit. À la différence des systèmes de dictée automatique par exemple, la limitation à un domaine d'application et à une forme particulière de dialogue permet de circonscrire les connaissances à mettre en œuvre dans les différents processus de traitement.

Nous proposons et justifions ici les bases d'une architecture permettant une forte intégration des différentes contraintes d'analyse. Notre ambition est de bénéficier de l'efficacité des modèles probabilistes et d'en limiter les biais pour l'interprétation en offrant la possibilité d'exploiter explicitement des connaissances expertes. Cette intégration repose sur l'homogénéité des structures manipulées par les niveaux de traitement, des automates synchronisés à partir desquels les différents modules créent des états temporaires de recherche, eux aussi synchronisés en conséquence. La manipulation de ces états temporaires est conditionnée par des contraintes propres aux mécanismes du niveau local de traitement (symboliques et/ou probabilistes) et par des contraintes globales impliquées par la synchronisation deux à deux des modules.

Ces différents principes sont actuellement expérimentés pour l'intégration d'un système de segmentation analytique de la parole continue, d'un module stochastique de reconnaissance de phonèmes et d'un analyseur basé sur les Grammaires Lexicalisées d'Arbres Adjoints (LTAG) synchrones pour les descriptions syntaxiques et de sémantique prédicative. Nos tests portent actuellement sur l'analyse de deux corpus de dialogue oral finalisé homme-machine obtenus par expérimentation de type Magicien d'Oz (corpus Wolff et GOCAD).

2. Les systèmes existants : leurs avantages et leurs limites

2.1. Architecture classique

Nous présentons en figure 1 l'architecture classiquement adoptée pour les composantes chargées de la reconnaissance et de l'analyse syntaxique des énoncés. Un HMM (*Hidden Markov*

1. Air Transport Information Systems. Le domaine d'application est celui de la recherche de renseignements sur les transports aériens dans une base de données.

Model) réalise le décodage acoustico-phonétique du signal. La combinatoire de reconnaissance est contrôlée dynamiquement par un modèle de langage de type *N-gram* (bi ou trigramme). L'intégration de ces deux traitements s'effectue exclusivement par la combinaison des probabilités. Le résultat de ce niveau de traitement infra-lexical est alors un treillis pondéré des hypothèses de mots, qui sera, dans un second temps, exploité par un module d'analyse syntaxique, le plus souvent grâce à une grammaire hors contexte ou à une analyse par segments (Abney, 1991).

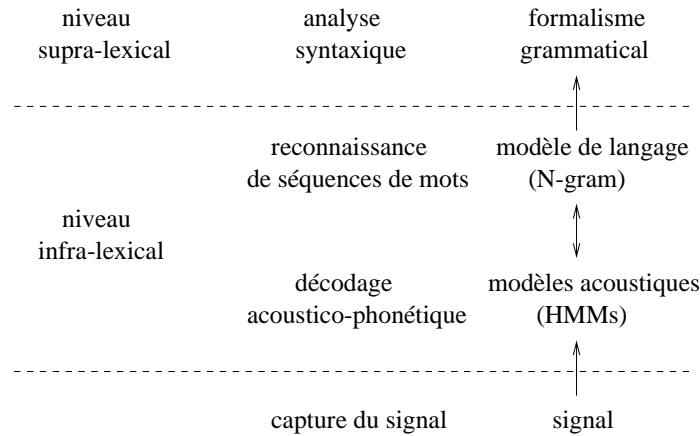


FIG. 1 – Composantes de reconnaissance et d'analyse syntaxique pour les architectures « classiques ».

2.2. Intégration par « treillis »

Une approche par treillis permet aux composantes syntaxique et sémantique de n'explorer que les meilleures hypothèses de reconnaissance, mais elle dissocie les composantes de reconnaissance et de structuration des énoncés, ce qui ne va pas sans poser un certain nombre de problèmes. Tout d'abord, le treillis est construit à partir des contraintes d'ordre des mots du modèle de langage. Cette technique est bien adaptée pour capturer des contraintes locales, mais est inefficace pour capturer des relations hiérarchiques, même simples, comme les accords sujet-verbe ou les rattachements prépositionnels particulièrement fréquents pour les dialogues de commande qui nous intéressent². De plus, les modèles de type *N-gram* acceptent des hypothèses d'énoncés agrammaticaux qu'on ne saura de toute façon pas interpréter et qu'il sera nécessaire de filtrer *a posteriori* (Roussel & Halber, 1997). Enfin le système de reconnaissance devant évaluer chaque mot à chaque point du signal comme une hypothèse séparée (Price *et al.*, 1989), cette approche est de surcroît coûteuse. Ces observations justifient l'intérêt d'une grammaire linguistiquement motivée qui permet notamment d'exprimer des contraintes de dépendance à distance. Il est important de noter que l'utilisation d'une telle grammaire linguistiquement motivée n'est pas incompatible avec celle de statistiques (grammaires stochastiques), offrant alors, en se substituant à un modèle *N-gram* de langage, des possibilités intéressantes de diminution de la *perplexité* de l'analyse. De bons résultats de perplexité ont été obtenus par exemple par (Srinivas, 1997) à l'aide d'un modèle de langage fondé sur un *supertagging* à partir d'une grammaire LTAG de l'anglais.

2. Notons néanmoins que ce résultat convient bien pour des approches où de toute façon l'interprétation sera obtenue par des méthodes sélectives sur des segments des hypothèses d'énoncés reconnus.

2.3. *Intégration « tout stochastique »*

Les bons résultats atteints par les systèmes stochastiques de reconnaissance (HMMs) sont dus à la rigueur de leurs fondements théoriques qui régissent les phases d'apprentissage et de reconnaissance. Cependant, il semble que le seul moyen pour améliorer l'aptitude des HMMs à capter la grande variabilité du signal soit d'augmenter continuellement le nombre de paramètres des modèles (utilisation de multigaussiennes, de modèles d'ordre 2, etc.) et parallèlement la taille des corpus indispensables à leur apprentissage. La stagnation constatée des performances des HMMs s'expliquent, à notre avis, par la grande difficulté rencontrée pour intégrer de manière explicite des connaissances acoustiques, liées à la production ou à la perception de parole. Quelques connaissances peuvent être prises en compte lorsqu'elles ne remettent pas en cause le cadre d'application des algorithmes d'apprentissage et de reconnaissance. Des informations perceptives simples ont pu ainsi être prises en compte en utilisant des techniques de paramétrisation de type PLP³, MFCC, qui utilisent une échelle perceptive non linéaire (Mel), ou issues d'un modèle auditif (Seneff, 1986). Des coefficients dérivés du premier et du second ordre sont ajoutés aux vecteurs de paramètres pour tenir compte de la forte corrélation acoustique observée sur certaines parties du signal et pour compenser les effets de l'hypothèse forte d'indépendance des observations successives. L'introduction de contraintes de durée des phonèmes est très complexe et se réduit généralement à adapter empiriquement la topologie des modèles au phonème considéré. La prise en compte de l'évolution dynamique de cette durée en fonction de la place du mot dans le syntagme et du rythme d'élocution est impossible à ce niveau. Malgré les performances atteintes par les HMMs, il semble que les contraintes structurelles (types de modèles, algorithmes) imposent *a priori* des limitations intrinsèques aux améliorations possibles.

La combinaison de probabilités constitue une solution relativement simple pour intégrer des informations hétérogènes. Se baser également sur un modèle de grammaire stochastique ou DOP⁴ pour mener l'analyse syntaxique permettrait alors une intégration directe des différentes contraintes statistiques des différents niveaux de traitement. Cependant la pertinence d'un énoncé ne peut se réduire à des critères statistiques et il nous semble nécessaire de compléter l'efficacité prédictive des trigrammes par des constructions résultant d'une expertise linguistique, capables d'apporter des éléments de décision sur des critères d'analyse autres que de simples considérations liées à la fréquence d'observation. Par d'exemple, (Roussel & Modave, 1998) proposent un modèle de décision multicritère flou pour déterminer le meilleur compromis entre critères de haut niveau et reconnaissance stochastique. Cependant cette évaluation s'applique, là encore, après analyse et ne permet que de classer les hypothèses produites. Enfin contrairement à ce que sous-entendrait une intégration stochastique, les rapports entre deux niveaux de traitement ne sont pas probabilistes, il s'agit d'informations statiques qu'il convient de représenter explicitement.

2.4. *Homogénéisation des connaissances et automates*

L'objectif d'intégration que nous souhaitons développer suppose une structure de représentation des connaissances commune à tous les modules hétérogènes. Nous considérons que les automates à états finis sont particulièrement bien adaptés pour les raisons suivantes :

- l'utilisation d'automates à états finis repose sur des techniques bien maîtrisées, simples à implanter et efficaces ;
- il est possible de représenter des structures complexes par des automates et de mener des

3. *Perceptually-based Linear Prediction* (Hermansky, 1990).

4. *Data Oriented Parsing*, voir (Bod, 1995)

recherches sur ces structures de données ;

- les automates pondérés permettent d'intégrer des considérations probabilistes (Pereira & Riley, 1991) ;
- les automates obtenus peuvent être minimalisés de façon à obtenir un partage optimal des sous-structures communes.

Notre hypothèse en particulier est que cette propriété de partage peut conduire à un taux important de factorisation des calculs d'analyse. Les automates à états finis sont utilisés couramment en reconnaissance de la parole : la représentation de mots sous la forme d'un graphe de phonèmes produit une structure compacte sur laquelle mener une reconnaissance. Les travaux du LADL ont par exemple démontré l'efficacité des techniques de minimalisation d'automates en lexicologie pour une tâche d'analyse et de génération de mots dérivés⁵. Nous allons maintenant présenter comment les différents niveaux de traitement peuvent être appréhendés comme des processus fonctionnant sur des données statiques représentées, malgré leur hétérogénéité, par des structures uniformes d'automates.

3. Automates de recherche et niveaux de traitement

3.1. Segmentation

3.1.1. Présentation du système

Le module de segmentation que nous utilisons repose sur la technique des réseaux de segmentation multiniveaux (Zue *et al.*, 1989) qui permettent de représenter l'ensemble des segmentations possibles du signal (de la plus fine à la plus grossière) dans une structure hiérarchique uniforme appelée *dendrogramme*.

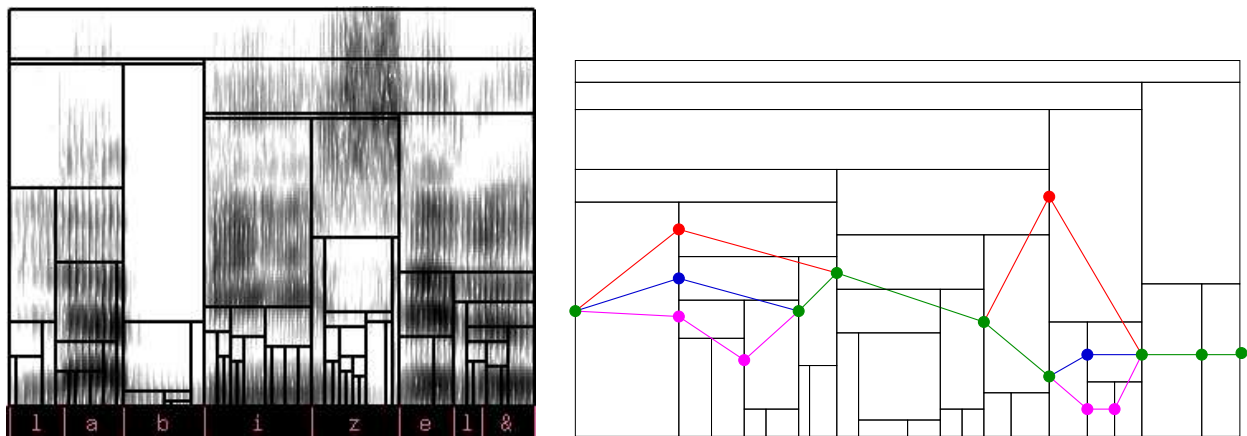


FIG. 2 – Exemples de dendrogramme réel et d'automate de solutions de segmentation sur un dendrogramme stylisé.

L'algorithme présenté dans (Husson & Laprie, 1996) permet de retrouver dans cette structure hautement combinatoire le sous-ensemble des solutions de segmentation phonétique les plus vraisemblables. L'efficacité de cet algorithme est due à l'intégration explicite de connaissances acoustiques, en particulier, un modèle de durée des phonèmes, un modèle d'homogénéité spectrale, ainsi que diverses contraintes telles que le respect des frontières de voisement

5. À titre d'exemple, l'automate qui permet la reconnaissance de 1.301.976 entrées n'occupe que 930 Ko en mémoire centrale et procède à hauteur de 15.000 mots à la seconde sur un ordinateur NeXT (Clémenceau, 1993).

détectées et l'estimation sous la forme d'un intervalle de confiance du nombre de segments acoustiques probablement présents sur le signal en fonction de sa durée observée et du mode de voisement associé. Ces critères et ces contraintes sont générés dynamiquement et combinés analytiquement pour fournir un ensemble réduit des n hypothèses les plus vraisemblables. Les résultats d'évaluation placent ce système en bonne place par rapport aux outils de segmentation phonétique concurrents (Husson, 1999).

Le système calcule incrémentalement durant le parcours du dendrogramme l'automate déterministe minimalisé représentant l'ensemble courant des segmentations les plus vraisemblables. Les états correspondent à des frontières temporelles entre des sons successifs. Les transitions sont étiquetées d'informations issues du processus de segmentation (durée calculée, macro-classe phonétique estimée). Un exemple d'automate minimalisé produit est fourni figure 2. Sur cet exemple, la minimalisation permet de passer de 30 à 16 états.

3.1.2. *Intérêts de l'utilisation de ce module*

Intéraction avec le module d'analyse syntaxique Les interactions avec le niveau de traitement supralexical s'effectuent par l'intermédiaire d'un automate phonétique lui-même contraint par un module d'analyse syntaxique utilisé en mode prédictif. Ce graphe phonétique permet d'affiner certaines contraintes exploitées par la segmentation comme le nombre de segments supposés présents, les séquences de classes phonétiques ou les transitions de voisement attendues. Réciproquement, le module syntaxique étant capable de situer le mot (et donc les sons correspondants) dans le groupe rythmique, ces informations peuvent également être propagées à l'automate phonétique et, donc, dans un second temps, utilisées pour ajuster les paramètres des modèles de durée utilisés par la segmentation. Parallèlement, les résultats de segmentation obtenus indépendamment de ces contraintes spécifiques permettent de classer voire de filtrer, sur la base d'indices acoustiques mesurés, certaines hypothèses d'analyse syntaxique.

Intéraction avec le module de reconnaissance stochastique Notre objectif n'est pas de contraindre la reconnaissance par HMMs par une segmentation *a priori* mais d'exploiter les hypothèses segmentales tout d'abord abductivement pour contrôler les modèles phonétiques appliqués à un moment du traitement de reconnaissance (selon la macro-classe phonétique ou le mode de voisement calculés pour la portion de signal acoustique considérée) mais aussi pour reclasser ou filtrer les résultats produits par le module de reconnaissance stochastique (confrontation de la segmentation temporelle, des transitions de voisement, les séquences de classes phonétiques obtenues par les HMMs et par le module de segmentation).

L'automate issu du processus de segmentation permet donc de mettre en œuvre des connaissances qui ne sont pas applicables directement dans une architecture « tout stochastique » et de réaliser une interface riche entre les niveaux de traitement infra et supralexicaux.

3.2. *Automate de partage pour grammaire lexicalisées d'arbres*

Une première technique de compaction de grammaires lexicalisées d'arbres par automates à états finis a été proposée dans (Evans & Weir, 1997). L'intérêt est de factoriser les sous-structures communes, qui sont extrêmement fréquentes dans ce style de grammaire, et d'éviter l'inconvénient majeur de la lexicalisation par rapport à une grammaire hors contexte classique, tout en préservant les avantages linguistiques et informatiques de grammaires de type LTAG. Dans la méthode proposée, la grammaire est initialement une LTAG sur laquelle on procède à une linéarisation de chaque arbre élémentaire. L'ensemble des arbres linéarisés peut être vu

comme un langage régulier dont les catégories forment l’alphabet. Pour chaque arbre élémentaire, nous construisons un automate déterministe acyclique permettant la reconnaissance de sa forme linéarisée, puis nous réalisons une minimalisation de l’ensemble de ces automates selon des techniques classiques. L’objectif est de partager les références utilisés par les états temporaires d’analyse sur cet automate, ceci afin de factoriser les opérations communes de l’analyseur.

L’automate utilisé ici n’impose pas de stratégie particulière d’analyse à la différence de ceux présentés dans (Evans & Weir, 1997). C’est sur cette structure précompilée d’automate qu’est maintenant mené l’algorithme ascendant d’analyse présenté dans (Lopez, 1998) sous la forme d’un ensemble de règles manipulant des *items* (structures temporaires indiquant une position sur l’automate) stockés dans un *chart*. Ces *items* représentent des îlots que l’on tente d’étendre de manière bidirectionnelle. Lorsqu’on ne parvient plus à étendre un segment grammaticalement, un deuxième ensemble de règles vient s’appliquer de façon monotone pour couvrir les phénomènes déviants de l’oral comme les reprises (Roussel & Lopez, 1999).

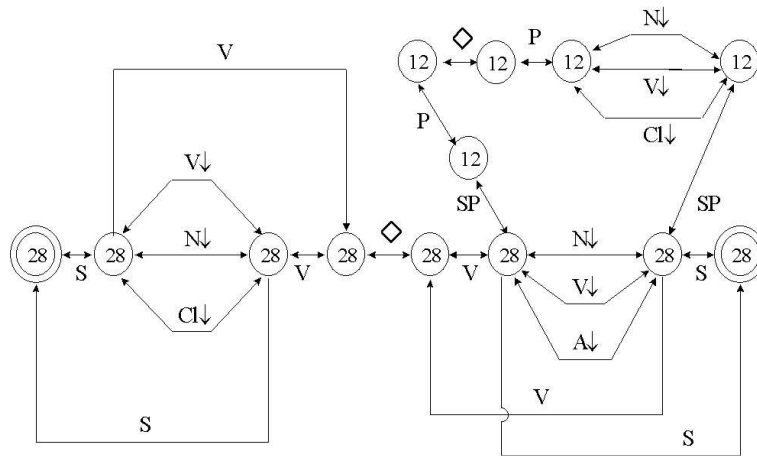


FIG. 3 – Automate représentant 28 arbres élémentaires correspondant à des contextes intransitifs et transistifs. Le symbole ◊ indique la position d’une ancre lexicale.

La figure 3 donne l’exemple d’un automate minimalisé permettant le partage de 28 arbres élémentaires. Le nombre dans les états indiquent combien d’arbres partagent cet état. On passe ici d’un nombre total de 273 états si on considère un automate par arbre (cas équivalent aux analyses classiques où chaque arbre élémentaire est considéré comme indépendant) à 13 états, donnant une idée sur le nombre d’opérations partagées pouvant avoir lieu durant l’analyse. Les évaluations récentes présentées dans (Carroll *et al.*, 1999) montrent sur le temps d’analyse un gain d’un facteur supérieur à 80 pour une grammaire LDTSG (Lexicalized D-Tree Substitution Grammar) à large couverture de l’anglais et des automates de type (Evans & Weir, 1997).

4. Intégrations d’hypothèses par automates

4.1. Synchronisation d’automates

L’idée de cascades d’automates à états finis et de combinaison de transducteurs pour la réalisation d’analyses sur plusieurs niveaux de traitement a été exploitée dans différents travaux, par exemple (Abney, 1991) et (Aït-Mokhtar & Chanod, 1997). La synchronisation quant à elle est une technique également largement éprouvée pour la traduction automatique. Les transducteurs d’états finis en permettant des traductions entre langages réguliers ont trouvé des appli-

cations notamment en analyses morphologique et phonologique. Les *PushDown transducers* (PDT) réalisent des traductions entre langages hors contextes. La synchronisation des Grammaires d'Arbres adjoints (TAG) a été proposée initialement dans (Shieber & Schabes, 1990).

La définition que nous utilisons ici de points de synchronisation entre deux automates ne repose pas sur la notion de transducteur, dans la mesure où nous ne souhaitons pas produire des traductions entre langages réguliers mais simplement établir des correspondances entre structures linéarisées se présentant sous la forme d'un langage régulier. De plus, notre objectif est d'une part d'éviter de dupliquer les structures au niveau des arcs de transitions, comme pour un transducteur, et d'autre part de ne pas combiner des automates qui relèvent de processus de traitement différents. La définition est la suivante : la *synchronisation* d'automates S est définie par le triplet (A_1, A_2, d_S) où A_1 et A_2 sont les automates synchronisés et d_S la fonction de synchronisation qui associe un ou plusieurs états de A_1 à un unique état de A_2 . Nous disons que la synchronisation entre A_1 et A_2 est assurée si pour chaque état σ_1 reconnu de A_1 son état synchronisé $\sigma_2 = d_S(\sigma_1)$ est également reconnu et que la réciproque est également vraie.

Au cours de l'analyse, en fonction des données à analyser, les modules de traitement initialiseront des états temporaires liés à au moins un état de ces automates. Par conséquent, la synchronisation est de fait étendue aux états temporaires d'analyse.

4.2. Synchronisation des traitements de reconnaissance

Le principe de synchronisation mis en œuvre peut être illustré grâce à la figure 4. Une hypothèse de segmentation correspond à un chemin dans l'automate segmental présenté figure 2. La fonction de synchronisation avec l'automate phonétique est construite dynamiquement avec l'automate segmental. On obtient alors un lien entre transitions correspondant à une macro-classe phonétique et les phonèmes possibles distribués dans le vocabulaire des corpus. Ainsi une hypothèse d'une segmentation à valider au niveau phonétique amènera à des hypothèses de mots, et inversement une hypothèse de mot suppose un certain chemin dans l'automate de segmentation. La même démarche est appliquée avec l'automate syntaxique : une hypothèse de mot correspond à un chemin dans l'automate de phonèmes et donc à une hypothèse d'ancre lexicale pour le niveau syntaxique.

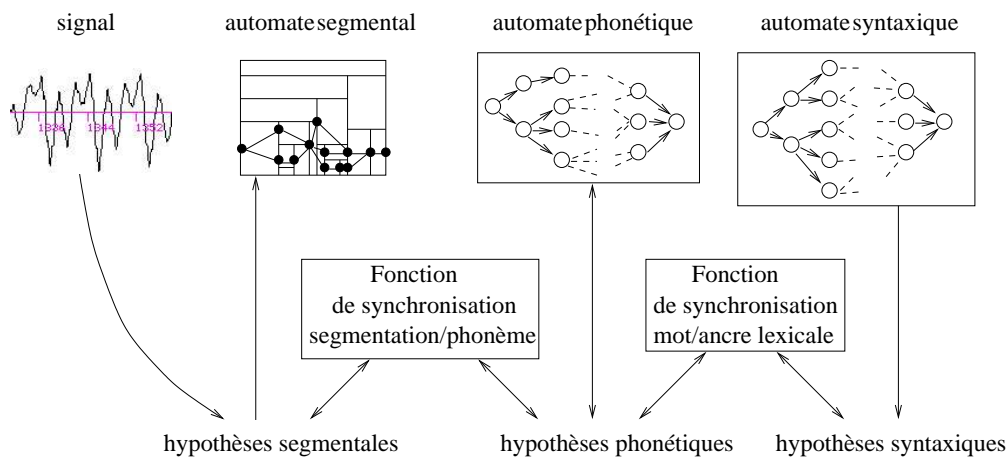


FIG. 4 – Principes de synchronisation des composantes RAP et analyse syntaxique.

Des structures temporaires sont créées par chaque module à partir de l'automate associé pour représenter les hypothèses concurrentes. L'indéterminisme de chaque niveau justifie notre stratégie de ne pas filtrer directement des hypothèses d'analyse indépendamment des autres

contraintes, comme c'est généralement le cas dans les systèmes existants. Aussi privilégions-nous les techniques tabulaires, couramment utilisées pour les traitements symboliques non-déterministes de l'écrit.

4.3. Génération des automates synchronisés

L'extraction des automates synchronisés est obtenue à l'aide du système EGAL (Lopez *et al.*, 1999), qui s'appuie lui-même sur le lexique BDLEX. Notre démarche est de réaliser un système dédié à une application et à un usage particulier de la langue. Les connaissances manipulées doivent donc être vues comme des spécialisations de connaissances générales de la langue afin de pouvoir les réutiliser au mieux d'une application à l'autre. La méthodologie associée suppose dans un premier temps un recueil de corpus par expérimentation de type Magicien d'Oz, que l'on supposera représentatif du sous-langage d'application, puis l'extraction automatique du vocabulaire employé. Cette liste de mot et les transcriptions phonétiques de BDLEX nous permettent de réaliser l'automate phonétique sur lequel se base le module de reconnaissance. Par ailleurs, de façon cette fois semi-automatique, nous extrayons la grammaire syntaxique de l'application dont nous synchronisons les reconnaissances des ancres sur celles, dans l'automate phonétique, des mots pouvant ancrer l'arbre élémentaire. Une première application de cette méthodologie sur le corpus GOCAD est reportée dans (Lopez *et al.*, 1999).

5. Liens avec les niveaux supérieurs de compréhension

(Shieber & Schabes, 1990) introduisent la notion de synchronicité entre grammaires TAG notamment pour l'interface entre syntaxe et sémantique. Les synchronisations entre arbres élémentaires d'une grammaire lexicalisée d'arbres syntaxiques et d'arbres sémantiques peuvent aussi se traduire dans deux automates minimalisés synchronisés⁶. D'autre part, un certain nombre de travaux récents tentent d'exploiter l'idée de grammaire lexicalisée d'arbres pour représenter la structure du discours, par exemple (Webber & Joshi, 1998). Les travaux réalisés ici pourraient alors être également étendus au niveau discursif.

6. Bilan

Nous avons proposé ici des principes généraux de synchronisation des données, en vue d'intégrer plus fortement les contraintes des différents modules de traitement des énoncés d'un système de dialogue oral. Dans la mise en œuvre effective d'un tel système, la question des contrôles des différents processus est fondamentale. L'avantage ici est que chaque niveau est susceptible de diriger une analyse, émettant des hypothèses valides selon le processus de traitement local, qui seront propagées aux autres niveaux de traitement chargés à leur tour de confirmer ou non ces hypothèses. Une architecture de type asynchrone, fondée sur un système multi-agents, nous semble le plus à même d'exploiter les propositions faites ici.

6. Notons ici cependant qu'en cas de reconnaissance partiel d'un automate, la synchronisation porte sur l'ensemble partiel des états synchrones du second automate qui peuvent ne pas être contiguës et donc faire perdre à ce deuxième niveau d'analyse le bénéfice de techniques tabulaires (analyse de complexité polynomiale au pire des cas devenant exponentielle au pire des cas). Cependant ce problème est général de l'interfaçage entre analyse syntaxique (généralement polynomiale, voir linéaire) et sémantique (potentiellement exponentielle car nécessitant une énumération des dérivations).

Références

- ABNEY S. (1991). *Parsing by chunks*. Kluwer Academic Publishers.
- AÏT-MOKHTAR S. & CHANOD J.-P. (1997). Incremental finite-state parsing. *ANLP'97, Washington*.
- BOD R. (1995). *Enriching Linguistics with Statistics : Performance Models of Natural Language*. PhD thesis, University of Amsterdam.
- BRISCOE E., GROVER C., BRANIMIR B. & CARROLL J. (1987). A formalism and environment for the development of a large grammar of english. In *10th International Joint Conference on Artificial Intelligence, Milan, Italy*, p. 703–708.
- CARROLL J., NICOLOV N., SHAUMYAN O., SMETS M. & WEIR D. (1999). Efficient parsing with wide-coverage lexicalized grammars. In *Submitted to ACL'99*.
- CLÉMENCEAU D. (1993). *Structuration du lexique et reconnaissance de mots dérivés*. PhD thesis, Université Paris 7.
- EVANS R. & WEIR D. (1997). Automaton-based Parsing for Lexicalized Grammars. In *Fifth International Workshop on Parsing Technologies, Cambridge, Mass.*
- HERMANSKY H. (1990). Perceptual linear predictive (PLP) analysis of speech. *Journal of the Acoustical Society of America*, **87**, 1738–1752.
- HUSSON J.-L. (1999). Evaluation of a segmentation framework based on multi-level lattices. In *to appear in EUROSPEECH, Budapest, Hungary*.
- HUSSON J.-L. & LAPRIE Y. (1996). A new search algorithm in segmentation lattices of speech signals. *ICSLP'96, Philadelphia, USA*.
- LOPEZ P. (1998). Analyse guidée par la connexité de TAG lexicalisées. In *Conférence sur le Traitement Automatique du Langage Naturel (TALN), Paris, France*.
- LOPEZ P., FAY-VARNIER C. & ROUSSANALY A. (1999). Sous-langages d'application et LTAG: le système EGAL. In *submitted to TALN'99, Cargèse*.
- LUZZATI D. (1989). *Recherche sur le dialogue homme-machine : modèles linguistiques et traitements automatiques*. PhD thesis, Université de la Sorbonne nouvelle Paris 3.
- PEREIRA F. C. N. & RILEY M. D. (1991). Speech recognition by composition of weighted finite automata. In *Proceedings of ACL'91, Berkeley, USA*.
- PRICE P., MOORE R., MURVEIT H., PEREIRA F., BERNSTEIN J. & DALRYMPLE M. (1989). The integration of speech and natural language in interactive spoken language systems. In *Proceeding of Eurospeech, Paris, France*.
- ROUSSEL D. & HALBER A. (1997). Filtering errors and repairing Linguistic Anomalies for Spoken Dialogue Systems. In *Workshop on Interactive Spoken Dialog Systems : ACL/EACL*, p. 74–81, Madrid.
- ROUSSEL D. & LOPEZ P. (1999). Contribution à l'analyse robuste non-déterministe pour les systèmes de dialogue parlé. In *submitted to TALN'99, Cargèse*.
- ROUSSEL D. & MODAVE F. (1998). A multicriteria scoring method to parse recognition hypotheses. In *the International Workshop on Speech and Computer (SPECOM), St.-Petersburg, Russia*.
- SELKIRK E. (1984). *Phonology and Syntax*. MIT Press, Cambridge.
- SENEFF S. (1986). A computational model for the peripheral auditory system: Application to speech recognition research. In *ICASSP '86*.
- SHIEBER S. & SCHABES Y. (1990). Synchronous Tree Adjoining Grammars. In *COLING*, volume 3, p. 253–260, Helsinki.
- SRINIVAS B. (1997). *Complexity of lexical descriptions and its relevance to partial parsing*. PhD thesis, University of Pennsylvania, Philadelphia.
- WEBBER B. L. & JOSHI A. K. (1998). Anchoring a Lexicalized Tree-Adjoining Grammar for Discourse. In *COLING, COLING-ACL'98 Workshop on Discourse Relations and Discourse Markers*.
- ZUE V. W., GLASS J. & SENEFF S. (1989). Acoustic segmentation and phonetic classification in the summit system. In *Proceedings Int. Conf. on Acoustics, Speech and Signal Processing*.