

IMPLEMENTATION DU SYSTEME MASPARG SELON UNE APPROCHE MULTI-AGENT

Chafik ALOULOU, Lamia HADRICH BELGUITH,
Ahmed HADJ KACEM, Souha HAMMAMI MEZGHANI
Faculté des Sciences Economiques et de Gestion de Sfax
Laboratoire LARIS, B.P. 1088, 3018 - Sfax - TUNISIE
Tél. (216) 74 27 87 77, Fax (216) 74 27 91 39
{chafik.aloulou; l.belguith; ahmed}@fsegs.rnu.tn; mezghani.souha@voila.fr

Résumé- Le traitement automatique du langage naturel est un axe de recherche qui connaît chaque jour de nouvelles théories et approches. Les systèmes d'analyse automatique qui sont fondés sur une approche séquentielle présentent plusieurs inconvénients. Afin de pallier ces limites, nous nous sommes intéressés à la réalisation d'un système d'analyse syntaxique de textes arabes basé sur l'approche multi-agent : MASPARG « Multi-Agent System for Parsing ARabic ».

Mots clés : Analyse syntaxique, système multi-agent, traitement automatique de l'arabe.

I. INTRODUCTION

L'objectif des traitements automatiques des langues est la conception de programmes capables de traiter de façon automatique des données linguistiques, c'est-à-dire des données exprimées dans une langue dite "naturelle".

Le processus du traitement automatique des données linguistiques nécessite plusieurs phases d'analyse : analyse morphologique, analyse syntaxique, analyse sémantique et analyse pragmatique [7].

Ces phases peuvent être couplées de différentes méthodes : séquentielle, parallèle, distribuée, etc.

Dans cet article, nous commençons par justifier l'apport de l'approche multi-agent pour l'analyse linguistique. Puis, nous présentons notre système d'analyse MASPARG (Multi-Agent System for Parsing ARabic) en exposant ses différents agents. Enfin, nous présentons un exemple d'analyse d'une phrase par le système MASPARG.

II. L'APPORT DES SYSTEMES MULTI-AGENTS

Les principes relevant du domaine des systèmes distribués permettent de résoudre les problèmes liés à une architecture séquentielle. En effet, un système multi-agent assure une coopération entre les différents niveaux de représentations, ce qui permet d'éviter de transmettre certaines solutions parasites d'un niveau à un autre.

Un système multi-agent permet aussi une décomposition des tâches complexes en sous-tâches, ce qui réduit la complexité du système au niveau du volume de connaissances et facilite le développement, le test et la mise à jour. De plus, une architecture distribuée assure une distribution de contrôle à l'aide des mécanismes de coopération et de coordination. Par ailleurs, un système multi-agent permet d'intégrer des stratégies nouvellement mises à jour pour le traitement de certains phénomènes linguistiques ou bien d'utiliser des méthodes différentes pour un même problème.

III. PRESENTATION DU SYSTEME MASPARG

Le système MASPARG est un système multi-agent composé d'un ensemble d'agents cognitifs, utilisant un contrôle décentralisé et une communication directe par envoi de messages. Ces agents collaborent ensemble dans le but de faire une analyse syntaxique d'une phrase donnée par l'utilisateur en déterminant son arbre syntaxique.

- **Agent Saisie :** L'agent « Saisie » est considéré comme une interface d'interaction entre notre analyseur et l'utilisateur puisqu'il lui permet de saisir son texte en langue arabe ou bien de choisir le fichier à analyser.
- **Agent Découpage :** Cet agent décompose un texte en paragraphes, lignes, syntagmes, groupement de mots, mots, etc. L'agent « Découpage » se base principalement sur des indicateurs de surface, à savoir les punctuations (; , . ? ! etc.), les numéros, etc.
- **Agent Lexical :** Cet agent propose une multitude de décomposition de mots ou pseudo-mots reconnus par l'agent « Découpage », en un ensemble de triades affixales (préfixe, affixe, suffixe) et de racines [5].
- **Agent Morphologie :** Cet agent a pour objet de reconnaître les caractéristiques morfo-syntaxiques du mot à savoir [3] : La partie du discours (verbe, nom, adjectif, préposition, etc.), le genre (féminin, masculin, etc.), le nombre (singulier, duel, pluriel), le mode, le temps et la personne avec laquelle il est conjugué (c'est le cas des verbes) L'agent « Morphologie » peut coopérer avec l'agent « Syntaxe » pour atteindre ses objectifs.
- **Agent Syntaxe :** Le but de l'agent « Syntaxe » est de trouver les bonnes analyses syntaxiques d'une phrase donnée. En fait, il détermine si une phrase appartient ou non au langage [4]. L'agent « Syntaxe » utilise les connaissances reçues de l'agent « Morphologie » et il fait aussi appel aux compétences d'autres agents, tel que, l'agent « Anaphore », l'agent « Ellipse » afin de réduire l'ambiguïté générée à ce niveau et éventuellement aux niveaux précédents.
- **Agent Anaphore :** Le but de l'agent « Anaphore » est de retrouver l'antécédent d'une forme anaphorique. Il établit aussi le lien entre la forme anaphorique et son antécédent. Le problème d'anaphore peut être particulièrement complexe et nécessite des connaissances syntaxiques plus approfondies, parfois même

supérieur au niveau syntaxique, à savoir le niveau sémantico-pragmatique [1].

- **Agent Ellipse** : L'agent « Ellipse » a pour but de reconstruire les différents types d'ellipses. Il utilise généralement des indicateurs et des règles de transformations afin de répondre aux demandes de coopération lancées par l'agent « Syntaxe » pour reconstruire une ellipse [2].

IV. IMPLEMENTATION DU SYSTEME MASPAP

Pour mettre en place le système MASPAP, nous avons utilisé la plate-forme AgentBuilder qui permet de définir les agents et les messages à échanger entre eux [6]. Aussi, nous avons développé les compétences des agents avec le langage JAVA.

Les connaissances nécessaires à l'analyse ont été représentées sous forme d'une base de données Microsoft Access. Pour la communication entre les différents agents, nous nous sommes alignés par rapport au langage de communication utilisé par AgentBuilder qui est le KQML (Knowledge Query Manipulation Language).

Nous nous sommes limités à l'implémentation de cinq agents seulement (i.e, agent Saisie, agent Découpage, agent Lexical, agent Morphologie et agent Syntaxe). Les agents Anaphore et Ellipse seront intégrés dans une étape ultérieure.

V. EXEMPLE D'ANALYSE D'UNE PHRASE PAR LE SYSTEME MASPAP

Considérons la phrase suivante : **رجع المهاجر إلى الوطن** (est revenu l'émigrant au le patrimoine.)*

L'agent Saisie envoie cette phrase à l'agent Découpage qui affichera le premier mot « **رجع** ». L'agent Découpage envoie le mot retrouvé à l'agent Lexical et continue à découper le reste de la phrase. Ainsi, l'agent Lexical commence à faire la décomposition du mot en affixes. Lorsqu'il trouve une décomposition valide, il l'envoie à l'agent Morphologie et continue à chercher d'autres décompositions valides.

La réception de la première décomposition valide, déclenche l'agent Morphologie qui détermine les caractéristiques morphologiques du mot reçu, l'envoi par la suite à l'agent Syntaxe puis refait le même travail pour les autres mots. L'agent Syntaxe, commence à sélectionner les règles de la grammaire contenant la même catégorie grammaticale que le premier mot « **رجع** ».

A partir de la liste des règles sélectionnées, l'agent Syntaxe renvoie à l'agent morphologie les caractéristiques possibles pour le deuxième mot. Ainsi, il demande à l'agent Morphologie si le deuxième mot « **مهاجر** » (émigrant) peut être ou non de type « **اسم فاعل** ». Puisque c'est le cas, l'agent Morphologie envoie un message de confirmation à l'agent Syntaxe.

Ces étapes se répètent pour chaque mot afin de raffiner la liste des règles possibles de la grammaire, pour la phrase en cours d'analyse, et d'orienter l'agent morphologie dans la détermination des caractéristiques morphologiques, ce qui permet d'éliminer les solutions parasites.

Lorsqu'il ne reste plus de mots à traiter, l'agent syntaxe affiche la structure arborescente trouvée.

VI. CONCLUSIONS ET PERSPECTIVES

Dans cet article, nous avons présenté l'analyseur syntaxique pour l'arabe basé sur l'approche multi-agent (MASPAP). Dans le but de prouver l'apport de cette approche, nous comptons faire une étude comparative du système MASPAP avec un système séquentiel de traitement automatique de la langue arabe (ce système est en cours de développement dans notre laboratoire de recherche).

Nous envisageons, comme perspectives de ce travail, d'intégrer d'autres agents à notre système, tels que, l'agent Anaphore ou encore l'agent Ellipse. Nous nous intéressons principalement à l'aspect communication avec les agents existant de MASPAP pour l'analyse des phrases complexes.

VII. BIBLIOGRAPHIE

- [1] Aloulou C., Hadrich Belguith L. et Ben Hamadou A. « Vers un système d'analyse syntaxique robuste pour l'Arabe : Application au recouvrement des erreurs de la reconnaissance », TALN '2000, Lausanne, Suisse, 2000.
- [2] Aloulou C., Hadrich Belguith L. et Ben Hamadou A. « Towards a more Efficient Linguistic Recovery of Handwriting Recognition », International Conference on Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications: ACIDCA '2000, Monastir, Tunisie, pp 31-36, 2000.
- [3] Belguith L., Ben Hamadou A. « Correction des erreurs d'accord basée sur une approche multicritère de classement des scénarios : le système DECORE », 11ème congrès sur la reconnaissance des formes et intelligence artificielle (RFIA '98), Clermont-ferrand, France, pp 271-280, vol. 3, 1998.
- [4] Belguith L. « Traitement des erreurs d'accord de l'arabe basé sur une analyse syntagmatique étendue pour la vérification et une analyse multicritère pour la correction », Thèse de doctorat en Informatique, Faculté des Sciences de Tunis, 1999.
- [5] Ben Hamadou A., « Vérification et correction automatique par analyse affixale des textes écrits en langage naturel : le cas de l'arabe non voyellé », Thèse d'Etat, Faculté des Sciences de Tunis, 1993.
- [6] Reticular System AgentBuilder, "An Integrated Toolkit for Constructing Intelligent Software Agents", San Diego 1999, 2000.
- [7] Stefanini M.H., « Talisman : Une architecture Multi-agent pour l'analyse du français écrit », thèse à l'université pierre Mendès-France, Grenoble, 1993.

* Ceci représente une traduction mot à mot.