Towards an Hybrid Approach
for Semantic Arabic Spontaneous Speech Analysis

Chahira Lhioui, Anis Zouaghi, and Mounir Zrigui,

ISIM of Medenine, Gabes University, Road Djerba, 4100 Medenine Tunisia,
ISSAT of Sousse, Sousse University, Taffala city (Ibn Khaldoun), 4003 Sousse,
FSM of Monastir, Monastir University, Avenue of the environnement 5019 Monastir,
LATICE Laboratory, ESSTT Tunis, Tunisia

Chahira_m1983@yahoo.fr, Anis.Zouaghi@gmail.com,
Mounir.Zrigui@fsm.rnu.tn,

Abstract. The automatic speech understanding aims to extract the useful meaning of the oral utterances. In this paper, we propose a hybrid original method for a robust automatic Arabic speech understanding. The proposed method combines two approaches usually used separately and not considered as complementary. This hybridization has the advantage of being robust while coping with irregularities of oral language such as the non-fixed order of words, self-corrections, repetitions, false departures which are called disfluencies. Through such a combination, we can also overcome structuring sentence complexities in Arabic language itself like the use of conditional, concession, emphatic, negation and elliptical forms. We provide, in this work a detailed description of our approach as well as results compared with several systems using different approaches separately. The observed error rates suggest that our combined approach can stand a comparison with concept spotters on larger application domains. We also present, our corpus, inspired from MEDIA and LUNA project corpora, collected with the Wizard of Oz method. This corpus deals with the touristic Arabic information and hotel reservation. The evaluation results of our hybrid spontaneous speech analysis method are very encouraging. Indeed, the obtained rate of F-Measure is 79.98%.

Keywords: Speech understanding, shallow parsing, 3-level HMM, Arabic speech processing, stochastic approach, cross-validation.

1   Introduction

The context of our work is the automatic Spoken Understanding Language (SLU) and finalized Human/Machine Communication (CHM).

Various kinds of linguistic knowledge are needed for the proper functioning of an SLU module. This linguistic knowledge can be of several types: lexical, syntactic, semantic, pragmatic and sometimes prosodic for systems taking as input a transcribed text. Many forms are offered to describe these linguistic knowledge kinds among

them we can mention: context-sensitive grammar [16], cases grammar [11], Hidden Markov Models [9], Neural Networks [28], N-gram language models [30], λ-calculus [32], logical [17] or Unification Grammars [1] in his various forms (UCG [19], Unification Categorial Grammar; APSG [31], Augmented Phrase Structure Grammar; LTAG [22], Lexicalised Tree Adjoining Grammar; STAG [26], Semantic Tree Association Grammar).

Moreover, we distinguish essentially in automatic SLU two types of approaches for the treatment of linguistic knowledge, which are linguistic approach [18] and stochastic approach [24]. In both cases, the statement is divided into word groups. These groups are frequently called concepts [9] [24] for stochastic approach and chunks [7] for a rule-based approach.

Whereas, these two approaches frequently used separately, enable a more or less effective understanding when dealing with oral speech. In fact, when a speaker speaks in a spontaneous way, the syntax or grammar errors are much more common in spoken than in written language. In one side, this problem is not guaranteed by a detailed linguistic approach. Besides, linguistic approach, as complete as they are, requires a grateful work to analyze corpus by experts in order to extract concept spotting and their predicates. This method is limited to specific fields using restrictive language. Thus, in the case of relatively opened field, it leads to many difficulties like portability and extension which are guaranteed by stochastic approaches.

In the other side, unexpected oral structures do not also obey to any statistical law and may not be satisfactorily modeled by a stochastic approach [23]. Besides, stochastic models do not seem to be able to solve the problem of a detailed language analysis. Added to that, the current stochastic models based on the only restricted linguistic entities observation sequences (words, syntactic categories, concepts, etc) cannot encounter statement deepness structures [23]. Finally, learning or training statistical techniques require a large amount of data annotated for learning stochastic model which is not always available.

However, it would be absurd to reject utterances which are syntactically incorrect because the goal is not to check the conformity of user's utterances to syntactic rules but to extract rather the semantic content. Hence, semantic level is important to control meaning of utterances and both syntax and semantic should have to be controlled in order to do not cause understanding problems. Stochastic models are more permissive than linguistic one based on formal grammars. They accept all the sentences of a language. Even incorrect sentences are accepted.

Qualities and drawbacks of these two separately-used approaches allowed us to detect certain complementarities between both of them. This observation is the motivation of our work that attempts to combine the two mentioned-above approaches to take advantage of their strengths. In this regard, we have proposed an hybrid approach based on linguistic and probabilistic methods to the semantic analyzer development for standard Arabic uttered sentence. This is achieved by integrating linguistic details describing local syntactic and semantic constraints on an Hidden Markov Model (HMM) stochastic model. The later materializes the language model of our Touristic Information and Hotel Reservations (TIHR) application.

2   Related Works and Motivation

Oral utterances are so often ungrammatical or incomplete and therefore an important part of the information contained in their texts is lost during an only syntactic rule-based analysis. That is why the analysis covering only syntax aspects is generally not effective. Added to that, Automatic Speech Recognition (RAP) systems generate a significant number of errors not contoured by grammars. Thus, to deal with all these oral treatment problems, some propose either a detailed linguistic phenomena analysis such as in [3], or a combination of a syntactic and semantic analysis such as in [31]. The others resort to stochastic methods as they are more robust to the transcription errors and adapt better to the specificities of oral language. Among these works, we cite for example the work of [9] which aims at achieving a stochastic conceptual decoding based on HMM with two levels in the context of robust spontaneous speech understanding. In the same context, [25] used also a Bayesian stochastic approach to speech semantic composition in Human/Machine dialogue systems.

Unlike Latin languages, the automatic understanding of spontaneous Arabic speech using stochastic approaches receives little attention at scientific research. In our knowledge, we note that the only work related to the use of stochastic techniques for the spontaneous Arabic language understanding are those of [35]. They have used a stochastic language model for spontaneous Arabic speech semantic analysis in the context of a restricted field (Train Information).

Thus, the originality of our work is to combine two common approaches used to be treated separately (linguistic and stochastic approaches) to the development of a semantic analyzer for standard Arabic utterances. This semantic analyzer is dedicated to tourists who communicate with the Interactive Voice Server (IVS) using Standard Arabic language in order to learn about touristic information that concerns them. Its principal role is then the construction of semantic representations for their utterances.

This hybrid approach consists, in fact, on integrating linguistic constraints in stochastic HMM model. Linguistic analysis should not inhibit the understanding process on the pretext that the input data are grammatically incorrect. According to the Blache citation appeared in [7], shallow parsing is a linguistic technique which facilitates greatly language processing. It rather conforms to spoken irregularities since it is only interested in extracting of word pieces containing only useful information (which are called chunks). Hence, a shallow parsing technique was chosen in our case.

Therefore, results of the linguistic analysis are syntaxico-semantic rules that governing chunks constituting user utterances. These syntaxico-semantic rules play the role of linguistic constraints applied in the first step of our hybrid analysis strategy. The second step will obviously be a stochastic HMM-decoding and then, HMM observations will be these extracted syntaxico-semantic constraints.

The choice to start with a linguistic analysis is appreciated from the fact that the latter does not allow if statements are utterly invalid. Hence, starting with a linguistic one prevents the spread of fatal errors between analyzer modules set in pipeline (see Fig. 1.).

We also note that our resulting hybrid model shows a difference from that of Bousquet [8]. This difference consists in having a 3-level HMM instead of a 2-level

Bousquet HMM. These three levels describe respectively three data types: syntaxico-semantic rules, conceptual and probabilistic information. Known that our application domain (TIHR) is relatively open, given its richness in terms of concepts, we numbers 83 between concepts and conceptual segment (see Section 4.3) and 163 linguistic rules (see Section 4.2).

3 Difficulties in Arabic Language speech Semantic Analysis

3.1 Disfluencies in spontaneous utterances

As indicates their Etymology, disfluencies mean any interruption or disturbance of influence [10]. In what follows, we focus on the major phenomena of disfluencies i.e. self-corrections, repetitions and hesitations.

— Self-correction: this case appears where the speaker made one or more mistakes and corrects in the same statement and in the real time. In this case, the wrong word (or words) is completely pronounced [8].
— Repeating: this is the case of the repetition of a word or series of words.
— Hesitation: it is a break filled in oral production which can be manifested in various ways: by using a specific morpheme (e.g. 'اَم' (Eum), 'اَه' (Euh) etc.) or in the form of a syllable elongation [10].

3.2 Semantic difficulties in Arabic language analysis

Arabic statements semantic analysis is a very difficult task given its semantic richness. This complexity is due to Arabic language specifics, which are:

— Arabic word can mean an entirely expression in English or French language [7]. For example the word "أرأيت؟" (ara ayta) designs in English language 'Did you see?". Thus, the automatic interpretation of such words requires a prior segmentation which is not an easy task [35].
— Words order in Arabic sentence is relatively variable compared, for example, to English or French languages where words order arrangement in a sentence respects perfectly the sequencing SVO (Subject Verb Object). Whereas, in Arabic, we always have the liberty to begin with terms in order to put the stress on. In addition, Arabic oral statements do not generally respect any grammar because of the oral spontaneous character. This complicates the task of building a grammar rules to be used in the semantic interpretation process. To do this, we must provide in such grammar all possible combination rules of words order inversions in a sentence.
— The non-capitalization of proper nouns, acronyms and abbreviations makes identification of such words even more difficult than the Latin languages.
— The connection without space of the coordination conjunction 'و' (and) to words. This makes it difficult to distinguish between 'و' as a word letter (e.g. وقف(stand)) and the 'و' having the role of a coordinating conjunction. But this type of combina-

tion plays an important role in the interpretation of a statement by identifying its proposals.

— The pronunciation nature of some Arabic letters, for example: غ(gh: ghayn pronunciation). These phonemes have no equivalent in other languages, such as, for example, French or English. Besides, some letters of the Arabic language, such as: ف(f: Fa pronunciation), ح(h: Hha pronunciation) خ(kh: Kha pronunciation) ض (d: Dad pronunciation), ذ (d: pronunciation Thal) ظ (z Zah pronunciation), are pronounced by a strong expiration; so the quality of the microphone can affect speech recognition results;

— The possibility of existence of many graphemes for the same phoneme (e.g. graphemes ض and ظ), or several realizations for the same phonetic grapheme (e.g. grapheme 'ل' (lam) has two different output sounds, depending on letters that precede and follow, as in: 'بالله' (please) and 'الله' (god)). Some graphemes may not be considered in pronunciation ('ا' for the elongation sound). This phenomenon makes difficult a speech recognition task and, consequently, the task of understanding.

4   Our Hybrid Method Approach

To carry out a robust Arab statements semantic analysis task, we decided to go through the two following steps:

— Chunking. This step is done by using a shallow rule-based parsing. Such analysis can often make a partial parsing in order to extract only essential elements that construct statements. These elements are called the chunks [13]. This analysis is guided in our case by a syntactic-semantic Probabilistic Context-Free Grammar (PCFG). The choice of opting for a PCFG is dictated by its adaptability to oral grammatical irregularities. In fact, in ASP (Automatic Speech Processing) field, the PCFG is often used in oral treatment cases. The result of chunking step is set of local linguistic constraints which govern extracted chunks.
— Stochastic analysis. During this step, a stochastic semantic decoding module transforms linguistic constraints extracted in the previous step on concepts.
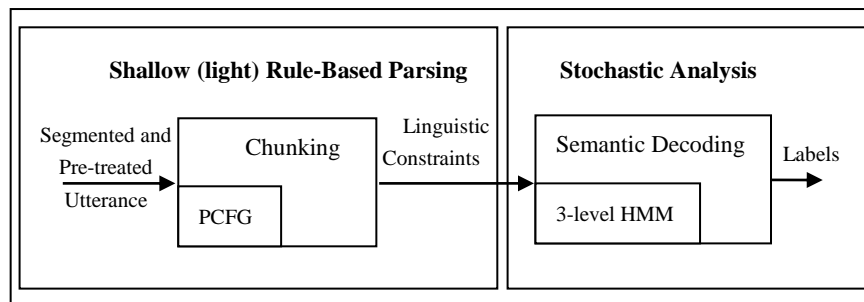

Fig. 1. The hybrid semantic analysis process

Fig. 1 describes our hybrid semantic analysis strategy adopted in Arab statements semantic analysis.

In the following, we detail each step separately.

4.1 Segmentation and Pretreatments

Because of the spontaneous utterance nature that contains various types of disfluencies (see at 3.)), an oral statement is inherently rigid and difficult to control. These disfluencies are frequent phenomena that appear normally in spontaneous speech. Here is an example of hesitation and self correction statement [12]:

<div dir="rtl">هل يوجد مطعم خاص بالكباب هنا آه عفوا بالبيتزا</div>

Is there a restaurant special kabab here, ah pizza sorry?

All these phenomena lead to the ambiguity problems. That is why a pretreatment step is required. This later removes duplication and unnecessary information, to divide complex utterances into elementary ones, to convert numbers written in all letters, and to determine the canonical forms of words [4] [14].

4.2 Chunking or Shallow rule-based Parsing

The basic idea of this analysis is to limit the depth and richness of a full parsing for the essentials. It consists on a partial parsing that promises to get minimal linguistic structures. These structures are called chunks. They are useful such as in other applications in that they promote their portability.

Our choice to use the chunk notion is based on the Blache citation which asserts that "chunks facilitate the analysis of a statement" [7]. To achieve this, we use the following chunk definition:

Definition1. A chunk is defined as a result of non-recursive categories formed of a head, to which function words and adjacent modifiers may be added [6].

According to the definition 1, we consider a chunk as typical segment consisting of a full word surrounded by functional words and adjacent modifiers. These are connected to the full word by the following linguistic rule:

$$Chunk\_Name \rightarrow NP$$

$$NP \rightarrow Det\ (Num)\ (Adj)\ Name$$

where the full word appears as a Name, the functional word is the Det and adjacent modifiers are (Num) and (Adj).

Shallow rule-based parsers are often characterized by a two-step process: pattern recognition step and word sense disambiguation one. Similarly, we envisage two independent steps for the realization of our shallow parser.

— Chunk identification
— Chunk attachment (through linguistic rules)

Chunk Identification. The main problem in a shallow rule-based parsing is the recognition of chunks. To resolve it, we applied a "Pattern recognition" algorithm used by many robust parsers [13] from which they take their robustness. The main idea of our algorithm is to loop the input string n times while calculating at each time the activation chunk degree present in this statement. Given that, according to what has been described in the cognitive ACT-R theory, it is possible to characterize a chunk by its activation level. This activation is described by the following equation extracted from [7]:

$$A_{ik} = B_i + \sum_j w_j \, S_{ji} \qquad \qquad \textbf{(1)}$$

where $k <= n$. In this formula, B represents the latency degree since the last access to the chunk. It is known as the basic activation and stores the frequency and the history of this chunk access. W corresponds, to the items or terms weight which are associated to the chunk. These weights are known as the 'sources' that can activate considered chunks.

Chunk Attachment. Relationships activating a chunk can be viewed as attachments properties that we seek to maximize. Solving chunk attachment problem is reduced to the study of two attachment types: intra-chunk attachment, which concerns words order inside the chunk, and inter-chunk attachments that defines relationships between different chunks constituting statements.

Intra-Chunk Attachment. The first attachment form affects words order inside a chunk. The example below shows a syntactic ambiguity of prepositional group attachment [with mayonnaise PP] either to the verbal group [want to eat VP] or to the nominal group [kebab NP]. This syntactic ambiguity causes a semantic ambiguity since we know more then, if the kebab is mixed with mayonnaise or the man asks two separated entities; one is the kebab and the other is the mayonnaise.

<div dir="rtl">أريد أكل الكباب بالمايونيز</div>
I want to eat kebab with mayonnaise.
(the kebab and mayonnaise are baked together) or (the kebab and the mayonnaise are baked separately)

Lexical ambiguities are compounded by the attachment ambiguities, especially syntactic ambiguities. In the dialog presented in the Fig 2, the initial user statement causes an analyzing system problem because, according to his knowledge, the lexical item "Richard" can be categorized either as a noun or as a surname. The rest of the utterance is analyzed without difficulty.

System: This is VIS system, hello, what is your request?

User: It is Richard, I would like to have some information about the X hotel reservation.

System: Is Richard a name?

User: Yes, it is.

System: Ok, what do you want exactly?

Fig. 2. A lexical ambiguity resolution

Regarding the resolution of lexical ambiguities, it can be assigned to the system in order to create an action. This action gives a communicative purpose that leads to the statement "Is 'Richard' a name?". The user confirmation of this hypothesis allows the system to finish the analysis of the initial user statement, and the search in the user's message box. Then we can say here that the interaction is closed by the user.

However, semantic ambiguities can occur even as in expressions where there are no syntactic or lexical ambiguities. For example, "the coast road" can be the road that follows the coast or the road that leads to it.

One example of syntaxico-semantic intra-chunk ambiguities are uses of condition constraints when the condition chunk 'لا.... إلا' (<not> ... <only>) is scattered on two fragments: الشرط and جواب الشرط where ' الشرط ' (the proposal) is the portion that lies between locution fragments <not> ... <only> and ' جواب الشرط ' (the proposal response) is the portion that comes after <only>. The example below illustrates what we are trying to explain:

–لا أريد السفر إلا على متن طائرة

-I do <u>not</u> want to travel <u>only</u> by a plane

where the use of the negation form in the proposal leads to a positive form in the proposal response to emphasize it. This complex formulation can easily and successfully be resolved by linguistic rules.

Inter-Chunk Attachment. The second form of chunks attachment concerns, in this case, the rearrangement of chunks within the same proposal. This problem increases with the Arabic language where words order is generally not fixed (see at 3.2). In fact, the sentence in the Arabic language do not undergo the SVO (Subject, Verb, Object) form, as it is the case of French or English languages. We cite two main categories which are the basis for the construction of Arabic[1] sentences: the first category is SVO (Subject Verb Object) that concerns nominal sentences, while the second, VSO (Verb Subject Object), is a form that describes verbal phrases. This is particularly critical in the case of Arabic language since it is morphologically rich.

– سوسة مدينة ساحلية(جملة إسمية)
- Sousse is a coastal city (Nominal sentence)
– أريد السفر إلى سوسة(جملة فعلية)
- I want to go to Sousse (Verbal sentence)

---

[1] The problem is the same for other source languages such as Chinese or Russian, and the solutions proposed here will therefore apply in other contexts.

Several tricks can also inhibit an automatic understanding of this language. Among them we can mention: reversals (تقديم وتأخير) of sentence constituents as it is the case of nominal sentences where the proposal (الخبر), putted as a prepositional phrase, may precede the theme (المبتدأ). The example below illustrates this reversal that results after the preposition placement of (حرف الجر) 'إلى' putted at the beginning of the subject:

<div align="center">

– أريد السفر إلى X

- I want to travel to X

– إلى Xأريد السفر

- To X I want to travel

</div>

A Constraints Grammar. A solution adapted to the intra and inter-chunk attachment problem is the description of different attachment constraints by a set of linguistic rules. This set is called "selection constraints". In this regard, we have designed and specified a probabilistic constraints grammar (see table below) with a probability distribution on the set of potentially finite rules defined by:

$$p(t) \geq 0$$
$$\sum_{c \in T_G} p(c) = 1$$

where t denotes the derivation tree which generates the constraint selection $c \in V^*$ and T_G designs all the derivation trees with G is a grammar describing all these constraints.

<div align="center">

Table 1. Definition of probabilistic constraints grammar

</div>

Sets	Description	Statistics
N : Set of non-terminal symbols	Syntactic, semantic and lexical terms : SN, SV, ADJ, VDestination, etc.	91
V : Set of terminal symbols	Linguistic constraints: e.g. C_El_Jar, cause/consequence, condition, WaElMaîya constraint.	43
S : Axiom	Chunk_attach	-
R : Set of production rules	The set of rules r_i linking the constraints	163
p$_i$: Probability attributed to the r_i rule	A positive coefficient attributed to each rule r_i	-

The elaboration of grammar constraints consists in the definition of:

− The inter-chunk syntax: this is equivalent to indicate how chunks can be arranged together. In our case, this information is modeled by bi-gram transitions that relies different chunks on each other.

− The intra-chunk syntax: this concerns words ordering within each chunk. Each chunk is also modeled by an HMM with bi-grams transition probabilities which relies all words that form the chunk.

For example, the linguistic selection constraint characterizing the condition phenomenon and connecting two chunks forming the condition rule is:

Chunk_condition → not C1 only C2
C1 → الشرط
C2 → جواب الشرط

Thus, the probabilistic specified constraints grammar can be represented as a 2-level HMM as described in the following figure:
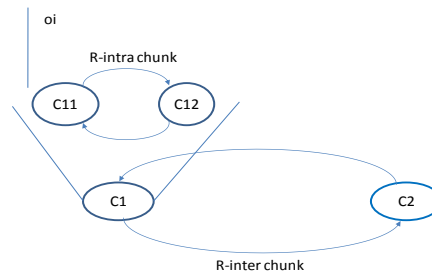

Fig. 3. The representation of constraint grammar on a 2-level HMM

4.3 Stochastic Analysis

Our stochastic analysis is based on concepts and conceptual segments notions. Their definitions are inspired from [9]. To achieve this, we used these two following definitions:

Definition 2. A concept (C) is defined as "a general and abstract mental representation of an object", and is independent of the language [8].

Definition 3. A conceptual segment (SC) comprises word sequences expressing the same meaning unit.

A word sequence making a conceptual segment is an instance of this conceptual segment. For example, the two following word sequences "at five o'clock" and "about four or five o'clock" are two instances of the Time conceptual segment.

According to definition 2. and definition 3., we conclude as a corollary 1. that a concept can be instantiated by several conceptual segments. In addition, according to the definition 1. and definition 3., we conclude as a corollary 2. that a conceptual segment can be formed by one or more disjoint chunks.

Example:

أريد[الذهاب /إلى سوسة]
I want [to travel \ to Sousse]

الذهاب (to travel) and إلى سوسة (to Sousse) chunks form the conceptual Travel segment.

Statement Semantic Representation Principles.

The user statement E is composed of n sequence words noted m_i.
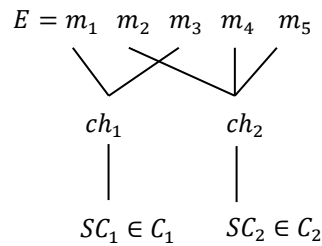
$$E = m_1\, m_2 \dots m_n$$

Then, we suppose the following hypothesis: Any statement E expresses a suite of k concepts. These concepts are also expressed using a suite of l conceptual segments and each conceptual segment is formed by a sequence of j chunks.

$$C = C_1\, C_2 \dots C_n$$
$$SC = SC_1\, SC_2 \dots SC_l$$
$$SC_i = ch_1\, ch_2 \dots ch_j$$

The statement E is linearly decomposed into series of l conceptual segments, and hence of $j \times l$ suites of chunks.

$$E = SC_1\, SC_2 \dots SC_l = (ch_j) \dots (ch_j)_l$$

For example, if the statement E has five words denoted m_i ($1 \le i \le 5$), and each m_i belongs to a chunk ch_k ($1 \le k \le 2$), and if E is divided into two conceptual segments SC_1 and SC_2 where each one belongs to a proper concept, a possible representation of the statement E is:

$$E = m_1 \quad m_2 \quad m_3 \quad m_4 \quad m_5$$

$$ch_1 \qquad ch_2$$

$$SC_1 \in C_1 \qquad SC_2 \in C_2$$

The principle of the statement interpretation with this representation is reduced then to divide into chunks, SC and concepts. This modeling with three levels of knowledge (chunks, SC and concepts) can be represented using a 3-level HMM.

Determination of Concepts and Conceptual Segments. The creation of a model language of a given application requires the determination of all concepts and SC that exist in this application domain. Detailed analysis of the corpus allows determining how these concepts are expressed in his different statements and what their corresponding SC refers to.

However, the relationship between concepts and SC is not always so obvious. In fact, many possibilities are offered to the application designer. Added to that, SC can overlap and belong to multiple concepts. We postulate that the meaning of a word depends on the SC where it is located. For example, consider the following two statements:

مدينة الانطلاق سوسة

<div align="center">

"Departing from Sousse"   (A)

في اتجاه سوسة

"On the way to Sousse"   (B)

</div>

The word سوسة (Sousse) exists in both statements but should be interpreted as a departure city in the statement (A) and as a destination city in the statement (B). So, according to the conceptual modeling principles of these statements, سوسة (Sousse) belongs to Departure conceptual segment in the statement (A) and to a Destination conceptual segment in the statement (B).

Another example that shows that the meaning of a word depends on the SC to which it is located. Consider the following statement:

<div align="center">

العاشرة و عشرة دقائق من فضلك

"A ten and ten minutes please"   (C)

</div>

The two numbers, in the statement (C), belong to the same concept Time and yet they must be interpreted differently. Indeed, the first number belongs to the Hour conceptual segment and the second belongs to Minute one.

Similarly, the interpretation of SC that compose the statement may depend on one other. Consider the following statement:

<div align="center">

لا أريد سوسة مدينة الانطلاق

الرفض   الانطلاق

"I do not want Sousse as a starting city" (D)

Refuse   Departure

</div>

The interpretation of chunks belonging to the Departure conceptual segment depends on chunks presence in the first SC.

Principles that we select for word interpretations in the context of the statement are:

— Word interpretation depends on the chunk where it is located;
— Chunk interpretation depends on the SC where it is located;
— SC interpretation of a SC depends on the concept from which it is issued;
— Word interpretation can be modified by the presence of previous or next chunks in the statement.

Semantic Analysis Principal. Semantic analysis principles are useful to interpret every word that composes statements according to chunks and SC to which they were assigned. The interpretation may eventually be modified later by the presence of previous or subsequent chunks.

The two tables below illustrate, in terms of concepts, SC and chunks, the conceptual domain of our application that deals with Tourist Information and Hotel Reservations (TIHR) (see at 5.1).

Table 2. Statistics of concepts, SC and linguistic constraints in our 3-level HMM

Concept (1-Level HMM)	Conceptual Segment (2-Level HMM)	Constraint rule (3-Level HMM)
3	80	163

Table 3. Identification of concepts, several SC and their corresponding chunks for the TIHR field

Concepts	Conceptual Segments (SC)	Chunks
Opening_Closing_Dialog	Request	{ « أريد » (I want), « أطلب » (I demand), « أرغب » (I desire), …}
	Accept	{ « أقبل » (I accept), …}
	Refuse	{ « أرفض » (I do not accept), …}
Hotel_Reservation	Hotel	{ « فندق » (hotel), « نزل » (hostel), …}
	Room	{ « بيت » (room), …}
	Tariff	{ « سعر » (price), « ثمن » (cost), « كلفة » (quote), …}
	Number	{ « عدد », …}
Touristic_Acknowledgement	Living_City	{ « قرية » (city), « بلدة » (town), « مدينة » (village), …}
	Itinerary	{ « طريق » (route), « الثنيا » (road), …}
	Transport_Mean	{ « وسيلة النقل » (mean of transport), …}
	Price_Ticket	{ « ثمن التذكرة », …}

Word interpretation depends on the SC where it is located follows the two principles below:

— Word interpretation depends on the concept that issued it;
— Word interpretation may be modified by the presence of a previous conceptual segment or according to the statement.

Semantic statement interpretation principles are useful to interpret every component in this statement according to word classes and SC to which it was assigned at the conceptual decomposition. The interpretation may possibly be changed later by the presence of previous or subsequent conceptual segments.

The 3-Level HMM Elaboration. The statement semantic analysis requires linguistic skills that are represented by a language model. According to our hybrid approach, our language model is defined in terms of concepts, SC and chunks attachment rules and hence modeled by a 3-level HMM. The observations of the model will be chunks simulated by linguistic constraints. These later were extracted by shallow parsing method based on PCFG analysis realized in the former step (see at 4.).

HMM Basic Principle. HMM models are powerful statistical tools that have been successfully used in various fields such as Automatic Speech Recognition (ASR) and Dialog Management (DM). They allow to model observation sequences putted in their two different forms; discreet and continuous form. Thus, we chose to represent our language model using an HMM model.

HMM are also N-grams language models. That is mean that the symbol emission probability depends on N-1 previous symbols. In our case, we consider that our HMM is a bi-gram language model.

The HMM model topology must be designed by a field expert of the considered application. Nevertheless, HMM parameters are automatically learned from training data. Therefore, stochastic models are more portable than linguistic ones where all rules must be explicitly written.

A 1-level HMM consists on two processes: the first is observable and the second is hidden. In all cases, following hidden states forms a Markov chain of order 1. A multi-level HMM is a model where observations of each hidden state are also modeled by a Markov model. The embedded Markov models number defines the model level number.

Retained Modeling. To be able to represent all the information that our language model carries out (linguistic rules, concepts and SC), a 1-level HMM is insufficient. We propose then a 3-level model (see Figure below) where: the first level is described by an HMM whose states represent our application concepts. According to the Table2. we have in total 3 principal concepts so, they correspond to three 1-level hidden states. Each concept is represented, in its turns, by an HMM whose states correspond to the conceptual segments (SC_i see Table 2.). Each SC will be represented by an HMM describing linguistic constraints of its realization.
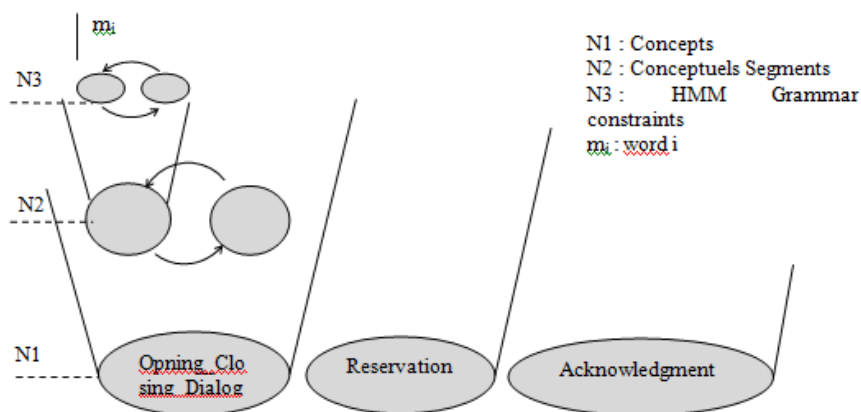
Fig 4. shows an overview of our 3-level designed HMM.

Fig. 4. Overview of our 3-level HMM

5   The Finalized Considered Application

To test our application and estimate the HMM parameters, we use a corpus dedicated to the study of touristic applications accessing to databases. The main reason for choosing this application field is the statistically representative size of the training corpus that we have. Moreover, through this corpus, we have the opportunity to produce an Arabic dialogue corpus in the same manner of those which are produced within the Francophone and Anglo-Saxon projects such as MEDIA [15] and LUNA [24]. Table 4. below shows a comparison between our Arabic dialog corpus and some other oral corpora issued in other fields and in other languages.

Table 4. Comparison between several International corpora and the our TIHR Corpus

Corpus	Language	Field	Size (on statement)
MEDIA	French	Touristic Reservation	18k
LUNA	Polonais	Transport Information	12k
TELDIR	Allemand	Time Train	22k
ATIS	English	Plan Ticket Reservation	6K
PlanRest	French	Restaurant Reservation	12k
TIHR Corpus	Arabic	TIHR	35k

5.1   Corpus Establishment

Oral corpora represent a significant proportion in the development of the automatic Spoken Language Understanding (SLU). These are closely linked to the availability of such corpus. However, Arabic speech corpora that deal with 'Touristic Information and Hotel Reservation' are very rare or even unavailable. Thus, to carry out our experiments in the context of this work, we were to build our own study corpus.

Our corpus is derived from the simulation of tourist information server and hotel reservations. Dialogues enunciated by tourists are in Standard Arabic Modern language (ASM). In fact, we consider that tourists are people who do not have Arabic as a native language. That is why their conversations were be uttered in the ASM language and not in common Arabic parlance.

These dialogues are about different themes such as: choice of living city, finding a route or a tourist event, a satisfaction of a price or date constraint. They were held between humans and machines through the Wizard of Oz protocol (Wizard of Oz, WoZ). Indeed, during the exchange, users believe converse with a machine while the dialogue is actually supported by a human operator that simulates informations and reservation server responses. The operator is assisted by the WOZ tool in the generation of responses to provide the user. After each user sentences, the operator shall consult the WOZ tool that offers to provide the answer to reflect the new dialogue state. To diversify the operator's responses, the WOZ tool is set at the messages, instructions and scenarios. A message sets are associated with the application to vary response formulations. With each call, the operator must follow instruction series (e.g. pretending not to have understood users to simulate mistakes which would make a

real system). These instructions must be provided to the WOZ tool and depend on the scenario chosen for the dialogue recording [12].

A detailed description of technical corpus characteristics is given in Table 5.

Table 5. Characteristics of our corpus (TIHR Corpus)

	Dialog Number	Speaker Number	Average Utterance Number / Dialog	Word Total Number
TIHR Corpus	10 000	1000	10	730 000

All queries corpus was recorded and transcribed manually as transcription standards in XML files, and labeled according to standards proposed by the ARPA community. Manual transcription was made a loyal way that it was recorded. That is to say, words are transcribed as we hear in recordings. Hence, we note disfluencies presence such as hesitations, self-corrections, repetitions, fault departures, etc.

In our corpus, we distinguish three query types i.e. independent query context, dependent query context, and absurd requests. The study of the corpus has allowed us to identify three concepts namely, Opening_Closing_Dialog, Hotel_Reservation and Touristic_Acknowledgement (see at Table 3.). In addition, we have defined SC among that we can mention: Living_City, Hotel, Travel, Tariff_Travel, Time_Travel, Price_Ticket, Period_Travel, Itinerary, Transport_Mean, Departure_City, Arrival_City, etc. Each conceptual segment is associated with a chunks set containing reference words relating to our application field. Words constituting chunks are linked by linguistic rules governing intra-and inter-chunk relations (see at Table 1.).

5.2 Tests and Evaluation

To avoid error propagation of the former linguistic analysis (surface analysis) to the second one (stochastic analysis), manual chunk verifications and their produced rules is performed just after the linguistic analysis to evaluate the shallow rule-based parsing results. An automation of this verification is suggested as one perspective. For the evaluation of the stochastic analysis, Kohavi cross-validation [28] is performed. In fact, this latter is used to validate an existing prediction or classification model or to find the best model by estimating its precision. It helps with a reduced bias to estimate the efficiency measure of our 3-level HMM. This measure is the average of each embedded HMM. The cross-validation technique is useful when the number of observations is fixed.

Cross-validation methods principle.

The cross-validation method principle is to divide the observation set into two separate and independent subsets where the first is called training set (that is the greater one) and the second is called validation or test set. The training set is used to generate the appropriate probabilistic HMM model. Nevertheless, the test set is used to evaluate the trained model according to the evaluation criteria.

Cross-Validation Initial Conditions. To perform cross-validation test, three initial conditions have to be satisfied. Indeed, we have to:

— prepare a sample observation
— have prediction model
— mesure the method performance for two separate sets either by calculating the rate error or by measuring the efficiency, precision, recall, F-score.

Cross-Validation Method Choices. Various cross-validation methods exist. Among them we can mention three principal ones that are respectively called: Holdout (also called Test set), Leave-one-out and k-fold. A detailed description of their advantages and disadvantages of each method is found in [5]. Therefore, we have not applied the Holdout method since it does not give good results because the observation set may be small for it. This badly affects the precision of performance values measured for model evaluation. In addition, the method of Leave-out-has been abandoned in our evaluation task because it is the slowest and the less used method. So we are simply satisfied to use just the K-fold method whose principle is to find the compromise of the two pre-cited ones.

K-Fold Method. Knowing that the parameter k often used in practice is $k = 5$ or $k = 10$ [21], the 5-fold and 10-fold are the two respective training methods which are used in our experiments. In both cases, we have divided the observation set in k disjoint partitions of equal size where $k - 1$ partitions are used for training and the left partition is used for the test. This process is repeated such as each partition is used one time for the evaluation. We then obtain k models and therefore, k accuracy measures. The final estimated accuracy value of the model learning from the training set is the average value of k calculated accuracies.

The Model Learning Step. In this step, the Baum-Welch learning algorithm uses the training observations and an initial model for generating a decoding model. Learning is to adjust the initial model parameters. The library used in our experiments is Jahmm (JAva HMM implementation) [37]. The latter allows the HMM implementation in Java and contains the basic algorithms for using HMM. To estimate the established HMM efficiency, we used the cross-validation method. This latter allows estimating the classification model effectiveness in general and for HMM models in particular.

Initial Hypothesis. Initial parameters choice for learning or training an HMM presents some difficulties [27]. However, in our case, we treat only discrete symbols. Thus, any observation probabilities distribution is applicable to our model.

Knowing that we have a 3-level HMM, where each level L_i ($1 \leq i \leq 3$) is characterized by n_i hidden states, we considered $1 + n_1 + n_2$ initial $HMM_{ij}(1 \leq j \leq n_i)$ which designate the ith level and the jth hidden state HMM. The characteristics of several models are shown in the following table. These models are classified as follow:

— HMM_1 for ($n_1 = 3$) concepts in the first level of the 3-level global HMM;
— $HMM_{21}, HMM_{22}, HMM_{23}$ for ($n_2 = 80$) conceptual segments in the second level. In this level we have ($n_1 = 3$) embedded HMM, where n_1 is the state number of the first global HMM level;
— ($n_2 = 80$) HMMs for $n_3 = 163$ linguistic constraints in the third level. This number is equal to the SC number that the second level contains.

Table 6. Initial models probability distribution features (for the first and the second global HMM level)

	HMM₁	**HMM₂₁**	**HMM₂₂**	**HMM₂₃**
π_i	Non-uniform	Uniform	Randomly	Non-uniform
A_{ij}	Non-uniform	Non-uniform	Randomly	Uniform
$B_i(o_t)$	Uniform	Uniform	Randomly	Non-uniform

The Learned Models. After the learning algorithm execution, the initial models parameters are gradually refined into a finite iteration number. We empirically chose 10 iterations. The latter value corresponds to learning algorithm stopping criterion. Following tables show average values of HMM initial probabilities simulated as two cases of k-fold cross-validation method applied for the 1-level HMM.

Table 7. Initial probabilities average value of the first level HMM with 5-fold cross-validation method

	π_1	π_2	π_3	π_4	π_5
HMM₁	0.2	0.3	0.3	0.15	0.05

Table 8. Initial probabilities average value of the first level HMM with 10-fold cross-validation method

	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
HMM₁	0.1	0.03	0.09	0.07	0.17	0.11	0.25	0.08	0.02	0.08

5.3 Evaluation

In order to calculate the performance criteria and thus to assess the decoded utterance quality, we generate an n-classes confusion-matrix (Table 3). It is obtained by comparing predicted labels sequences (resulting states of the decoding step) and real labels sequence (extracted manually). Real labels sequence is determined by confusion-matrix rows information. As for predicted labels, they correspond to the confusion-matrix columns information (see Table 9).

Diagonal confusion-matrix C_i^j cells represent correct predicted labels. They correspond to the C_i label occurrence number in two state sequences. Besides, confusion-matrix C_i^j cells represent incorrect predicted labels with $i \neq j$. They correspond to real and predicted C_i and C_j label occurrence numbers.

Result Interpretation. To evaluate the learning generated models, we have produced the most likely predicted states sequence (generated by the learned HMM) associated to a given observation sequence (sequence of real states). This is done for all test set observation sequences.

Subsequently, we built confusion matrices from predicted and real state sequences according to the chosen cross-validation method. Specifically, we have obtained five confusion-matrices using the K=5 method and ten confusion-matrices using K=10.

To illustrate this, we present a confusion-matrix for three labels representing the 1-level HMM states. This matrix has been obtained with the method K=5 on its first test set.

Found: L1 L3L1L2L1L3L3L2L3L3
Predicted: L1L3L1L1L1L3L3L3L3L3

Table 9. Example of a confusion matrix to 3 classes

		Prédites		
		L1	**L2**	**L3**
Réelles	**L1**	**3**	0	0
	L2	1	**0**	1
	L3	0	0	**5**

In the above-shown confusion-matrix, all label L1 predictions are correct. They are equal to three. For label L2, there is no correct prediction. On the other side, for label L3, five predicted labels are correct. We have in total eight correct labels among ten ones. We conclude then that the model generates a satisfying correct prediction numbers. Therefore, the decoding quality is satisfactory.

Performance Criteria. Table 10 presents precision, recall, and F-score average values for HMM_1, HMM_{21}, HMM_{22} and HMM_{23} embedded trained models of our 3-level HMM using K5, K10 methods. These metrics were calculated from confusion matrices for each test set.

Table 10. Precision, Recall and F-score average values for some embedded trained models using K5, K10 methods.

Modèles	Précision	Rappel	F-score
HMM_1	70.00%	71.00 %	73.79%
HMM_{21}	71.08%	68.99%	74.1%
HMM_{22}	69.98%	70.89%	72.9%
HMM_{23}	70.02%	72%	73.77%

Results Comparison. Comparison results between the Bousquet's CACAO stochastic conceptual decoder [9] and the Zouaghi and Zrigui's semantic decoder [36] shows that the error response rate, obtained by our system, was reached 20.02% which is considerably less than Bousquet system (see Table 11).

Table 11. Comparison of our hybrid system results with several well-known systems

	CACAO	Oréodule Project decoder	Our analyzer
Approach type	stochastic	semantic	Hybrid
% Error rate	29.11%	29%	20.02%

6   Conclusion and Perspectives

One of most supervised objectives that we hope to achieve when we implemented our semantic analyzer system was to fulfill spontaneous spoken Arabic language robust parsing while making the application field wider than it is currently done. Linguistic approaches are not usually viewed as efficient tools for pragmatic applications. That's why we were interesting in combining two frequently separately-used approaches (linguistic and stochastic approaches).

A second objective was to have rather a generic system, despite a field-based linguistic knowledge use. This constraint is achieved through generic rule definitions as well as their probabilities for the third linguistic integrated HMM level training. This makes it possible to estimate efficiently its parameters. Our analyzer performances show that the two divergent approaches combination can bear comparison with systems which are based on a lonely approach (stochastic approach for CACAO Bousquet's conceptual decoder and semantic approach for Zouaghi and Zrigui semantic decoder).

The model was trained using the Baum-Welch algorithm. This adjusts the initial model parameters. The learning step was carried out according to different cross-validation techniques. Then, the model evaluation was also carried out in the quantitatively and well-known methods using different metrics such as precision, recall and F-score.

As a perspective, we hope to refine our system analysis to be tested in a second challenge by evaluation campaign that will focus on the phenomena described in the typology that we have proposed.

References

1. Anne Abeillé, *Les Nouvelles Syntaxes: grammaires d'unification et analyse du français* (Linguistique). Paris: Armand Colin, 326 pp. 2 200 21096 5, 1993
2. Antoine Jean-Yves, Goulian Jérôme, Villaneau Jeane, Quand le TAL robuste s'attaque au TAL parlé : analyse incrémentale pour la compréhension de la parole spontané, TALN 2003, Batz-sur-Mer, 11-14 Juin 2003
3. Antoine, J.-Y., "Pour une Ingénierie des Langues plus Linguistique", *HDR Computer Science*, University of South Bretagne, Vannes, France, 2003
4. Bahou, Y., Belguith, H.L., Ben Hamadou, A.: Towards a Human-Machine Spoken Dialogue in Arabic. In: 6th Language Resources and Evaluation Conference (LREC

2008),Workshop HLT Within the Arabic World. Arabic Language and Local Languages Processing Status Updates and Prospects, Marrakech, Morocco, 2008b

5. Besbes G., Modélisation De Dialogues A L'aide D'un Modèle Markovien Caché Mémoire Présenté, Mémoire présenté à la Faculté des études supérieures de l'Université Laval , 2010

6. Bird S., Klein E. et Loper E., Natural Language Processing with Python, O'Reilly Media, 2009

7. Blache Philippe, Chunks et activation, un modèle de facilitation du traitement linguistique, TALN-RÉCITAL, 17-21 Juin, Les Sables d'Olonne ,2013

8. Bouraoui J.-L., Traitement automatique de dysfluences dans un corpus linguistiquement contraint, *Actes de la 16ème Conférence sur le Traitement Automatique des Langues Naturelles*, *TALN'09,* Senlis, France, 2009

9. Bousquet-Vernhettes C., Compréhension robuste de la parole spontanée -Décodage conceptuel stochastique, Thèse de doctorat, Université Paul Sabatier, 2002

10. Bove R. A Tagged Corpus-Based Study for Repeats and Self Repairs Detection in French Transcribed Speech, Proceedings of the 11 th International Conference on Text, Speech and Dialogue, TSD'08, Brno, Czech Republic., 2008

11. Bruce, B., "Cases Systems for Natural Languages", *Artificial Intelligence*, Volume 6, pp. 327-360, (1975)

12. Chahira L., Anis Z., Mounir Z., A combined A Combined Method Based on Stochastic and Linguistic Paradigm for the Understanding of Arabic Spontaneous Utterances, A. Gelbukh (Ed.): CICLing 2013, Part II, LNCS 7817, pp. 549–558, 2013.

13. François Trouilleux, Un analyseur de surface non déterministe pour le français, TALN 2009, Senlis, 24–26 juin 2009

14. Hadrich Belguith, L., Bahou, Y., Ben Hamadou, A.: Une méthode guidée par la sémantique pour la compréhension automatique des énoncés oraux arabes. International Journal of Information Sciences for Decision Making (ISDM), Septembre 2009

15. H.BonneauMaynnard, K. Mctait, D. Mostefa, L. Devillers, S. Rosset, P.Paroubek, C. Bousquet, K. Choukri, J. Goulian, JY Antoine, F. Béchet, O. Bontron, L. Charnay, L. Romary, N.Vergnes, N. Vigourous, Constitution d'un corpus de dialogue oral pour l'évaluation automatique de la compréhension hors et en contexte du dialogue, Actes des Journées d'Etude sur la Parole (JEP 2004), Fès, Maroc, avril 2004.

16. Jean-Michel Autebert, Jean Berstel et Luc Boasson, *« Context-free languages and pushdown automata »*, dans G. Rozenberg, A. Salomaa (éditeurs), *Handbook of Formal Languages*, vol. 1 : *Word, Language, Grammar*, Springer Verlag, 1997 (ISBN 978-3540604204)

17. Jeanne Villaneau, Olivier Ridoux, Jean-Yves Antoine, Comprehension de l'oral spontané Présentation et évaluation des bases formelles de LOGUS RSTI RIA.Volume18– n° 5-6/2004, pages 709 à 742, 2004

18. J. Allen., Natural language understanding. Redwood City, CA, USA : Benjamin-Cummings Publishing Co., Inc. 28, 1988

19. Karine Kray-Baschung, Gabriel G. Bes, Thierry Guillotin, French Unification Categorial Grammars https://halshs.archives-ouvertes.fr/halshs 00371433 , 2014

20. Kurdi, M-Z., Contribution à l'analyse du langage oral spontané, Thèse de l'Université Joseph Fourier, Grenoble, 2003

21. Liu, B., Web Data Mining - Exploring Hyperlinks, Contents and Usage Data, Springer, 2006.

22. Lopez P., "Analyse guidée par la connexité de TAG lexicalisées", TALN'98, Paris, 1998

23. Marcel Cori, Des Méthodes De Traitement Automatique Aux Linguistiques Fondées Sur Les Corpus, Langages, n° 171, p. 95-110. DOI : 10.3917/lang.171.0095, 2008.

24. Marie-Jean Meurs, Approche stochastique bayésienne de la composition sémantique pour les modules de compréhension automatique de la parole dans les systèmes de dialogue homme-machine, THÈSE présentée à l'Université d'Avignon et des Pays de Vaucluse, 2009

25. Meurs M., Lefèvre F. et De Mori R. , Spoken Language Interpretation: On the Use of Dynamic Bayesian Networks for Semantic Composition, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2009

26. Mohamed-Zakaria Kurdi, « A spoken language understanding approach which combines the parsing robustness with the interpretation deepness », International Conference on Artificial Intelligence (IC-AI'01), Las-Vegas, États-Unis, 2001

27. Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, February 1989.

28. R. Kohavir., A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, San Francisco, CA, USA, pp. 1137-1143, 1995

29. Salma Jamousi, Kamel Smaili et Jean-Paul Haton, « Contribution à la compréhension de la parole par des réseaux neuronaux », Quatrième Rencontres Jeunes Chercheurs en Parole (RJC'01), Mons, Belgique, p. 70-73, 2001.

30. S. Knight, G. Gorell, M. Rayner, D. Milward, R. Koeling, I. Lewin. Comparing grammar-based and robust approaches to speech understanding: a case study. *Proceedings of European conference on speech communication and technology*, 2001.

31. Stephanie Seneff, « TINA: A Natural Language System for Spoken Language Applications », Computational Linguistic, p. 61-86, 1992.

32. Villaneau, J., Antoine, J.-Y., and Ridoux, O., Logical approach to natural language understanding in a spoken dialogue system. In Sojka, P., Kopecek, I., and Pala, K., editors, The International Conference on Text, Speech and Dialogue (TSD), volume 3206 of Lecture Notes in Computer Science, pages 637–644. Springer, 2004

33. Tom Brondsted, « The Linguistic Components of the REWARD Dialogue Creation Environment and Run Time System », IEEE 4th Workshop Interactive Voice Technology for Telecommunications Applications (IVTTA'98), Turin, Italie, p. 71-74, 1998.

34. Zouaghi, A., Zrigui, M., Antoniadis, G.: Compréhension Automatique de la Parole Arabe Spontanée. Traitement Automatique des Langues (TAL 2008) 49(1), 141–166, 2008

35. Zouaghi, A., Zrigui, M., Antoniadis, G., Présentation d'un modèle numérique pour la compréhension de la parole arabe spontanée, 2007

36. Zouaghi, A., Zrigui, M., Considération du contexte pertinent pour améliorer les performances d'un étiqueteur sémantique de la parole arabe spontanée, RJC, 2005

37. http://www.run.montefiore.ulg.ac.be/~francois/software/jahmm/