Statistical Framework with Knowledge Base Integration for Robust Speech Understanding of the Tunisian Dialect

M. Graja, M. Jaoua, and L. Hadrich Belguith

***Abstract*—In this paper, we propose a hybrid method for the spoken Tunisian dialect understanding within a limited task. This method couples a discriminative statistical method with a domain ontology. The statistical method is based on conditional random field (CRF) models learned from a little size corpus to perform conceptual labeling task. These models are able to detect the semantic dependency between words. However, the domain ontology is used to add prior knowledge about the task. Our experiments are based on a real spoken Tunisian dialect corpus. The obtained results show that the proposed method is able to improve the performance of CRF models for speech understanding by the integration of the domain ontology. Our method can be exploited for under-resourced languages and Arabic dialects to overcome the lack of linguistic resources.**

***Index Terms*—Conditional random field (CRF), domain ontology, knowledge base, speech understanding, statistical models, Tunisian dialect (TD).**

I. Introduction

SPOKEN Language Understanding is a crucial component in spoken dialogue systems. It aims to clarify the meanings from spontaneous speech [1]. The first level of spoken language understanding is concept labeling, which consists in the segmentation and extraction of semantic concepts from transcribed speech. Indeed, the concept labeling task takes the transcribed words as input and provides conceptual labels as output.

To perform conceptual labeling, two approaches have been proposed. The first one includes knowledge-based methods which have been widely used [2][3]. They consist in writing rules to model the structures of utterances. They require much expertise to define rules and verify them manually. The second one consists of statistical models which can be classified into generative and discriminative models [4]. These statistical models guarantee dependencies between the words and gather a rich set of features to facilitate conceptual labeling [5]. But they require large spontaneous speech corpora [6]. These models rely on parameters estimated from annotated data. In the case of speech understanding, annotated data are turns which are semantically labeled. So, annotation schemes based on conceptual labels must be defined to annotate utterances and then, find the statistical model parameters. Finally, these parameters are used to infer the statistical models and label new examples of transcribed speech.

Statistical methods have been widely applied to speech understanding. Martínez-Hinarejos *et al.* used a statistical method based on Hidden Markov Models (HMM) for the Spanish language, using the DIHANA corpus [7]. The DIHANA task deals with requests of information about railway services [8]. Their work consists in using HMM in the most realistic situations where dialogues are not segmented into utterances. The results of their work are interesting since they obtained 92% as F-measure. This is due to the large size of the corpus used for training models. However, several works have shown the robustness of Conditional Random Fields (CRF) models for request information in the French language; using the MEDIA corpus [4][9]. The MEDIA corpus is manually annotated with semantic concepts according to touristic information [9]. But it should be noted that the turns in this corpus are segmented into utterances, which facilitates the understanding task. Raymond *et al.* [4] has used CRF models and then incorporated the domain knowledge through a set of rules made manually. This has reduced the conceptual error rate (from 11.2% to 10.9% as CER), and has increased the performance of the system to 92% as F-measure. This justifies the advantage of segmenting turns into utterances and the important size of the training corpus.

Many works have proved that the CRF models are the best among generative and discriminative models [10]. These models give excellent results for many tasks in Natural Language Processing, even for speech understanding [4]–[11]. However, these models have generally been applied to conceptual labeling of spontaneous speech in Latin languages such as English, Spanish or French [12], and not for Arabic dialects.

The Arabic dialect is the form of language used in daily conversations. Despite this fact, the works which dealt with the Arabic speech understanding problem were interested only in MSA (Modern Standard Arabic) [13][14] and not in the Arabic dialect. However, a dialogue system in MSA is not an interesting

Manuscript received October 29, 2014; revised April 19, 2015; accepted July 27, 2015. Date of publication August 18, 2015; date of current version September 22, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. James Glass.

M. Graja is with the ANLP-RG, Miracl Laboratory (Multimedia, InfoRmation Systems and Advanced Computing Laboratory), University of Sfax, Sfax 3072, Tunisia (e-mail: marwa.graja@fsegs.rnu.tn).

M. Jaoua and L. H. Belguith are with the ANLP-RG, Miracl Laboratory (Multimedia, InfoRmation Systems and Advanced Computing Laboratory) and the Department of Computer Science, Faculty of Economics and Management of Sfax (FSEGS), Sfax 3018, Tunisia (e-mail: maher.jaoua@fsegs.rnu.tn; l.belguith@fsegs.rnu.tn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2015.2464687

application in speech technology since it is difficult to Arab speakers to reformulate a correct utterance in standard Arabic. In addition, the Arabic dialect is used to fulfill all public services and it is the most useful language form for daily conversations. Therefore, it is so crucial to consider dialects in dialogue systems.

In this paper, we propose to evaluate three methods for the Tunisian dialect (TD) understanding. The first one is a knowledge-based method which integrates domain ontology. In the second method, we propose to evaluate a discriminative model to conceptually label the transcribed speech in the TD. Indeed, we are interested in evaluating a discriminative algorithm based on CRF models and learned from a small-size corpus with different levels of processing data. In the third method, we couple the two methods into a hybrid one to ameliorate the conceptual labeling of the spoken TD.

Corpora in the TD are very rare and sometimes undergo development. In this paper, the corpus used for testing the proposed methods is the TUDICOI corpus. It is a task-oriented spontaneous speech dialogue corpus for a railway inquiry task in the TD.

In this paper, we present two major contributions. Our first contribution consists in modeling CRF models through non-segmented turns. This allows us to reduce the manual segmentation of turns into utterances. Our second contribution is to propose a method for automatic conceptual labeling for the TD as an under-resourced language. We learn CRF models from a very small-size corpus, since the acquisition of a corpus is a difficult task. Then, we integrate a domain ontology to improve the labeling task by adding new knowledge to full cover the used lexicon. This method is very interesting for Arabic dialects which suffer from a lack of linguistic resources.

This paper is organized as follows. Section II presents an overview of the TD at different levels. Section III describes a corpus for the spoken TD used for experiments. Section IV presents an understanding method based only on a domain ontology. Section V deals with statistical models based on CRF for conceptual labeling. We present in Section VI a hybrid method which consists in coupling the CRF models with the domain ontology. Experiments and evaluation of each method are presented in Section VII. A conclusion is drawn in the last section.

II. The Tunisian Dialect

The Arabic language consists of three main clusters: Classical Arabic, Modern Standard Arabic (MSA) and Dialectal Arabic. The latter is the informal form of the language since it is used for daily conversations and for requesting information [15]. That is why we should consider dialects in the spoken dialogue systems. The Tunisian dialect (TD) is a subset of Arabic dialects related to the Arab Maghreb (western Arab world). The TD is strongly influenced by other languages such as Turkish, Italian and French. It represents a mosaic of languages. It has several large regional varieties, but the variety of Tunis (used in the capital of Tunisia) is the most understood by all Tunisians [16].

Like all Arabic dialects, the TD is characterized by a morphology, a syntax, a phonology, a lexicon and an orthography that have similarities and differences in comparison with the MSA and even with other Arabic dialects. In addition to that, the TD's processing is a complicated task for several reasons. This is mainly due to the lack of linguistic resources and processing tools such as transcribed texts, annotated corpora, dictionaries and morphological and syntactic analyzers.

Below we present some characteristics of the TD at different levels.

A. Phonological Level

The TD is divided into six major dialectal areas as follows: the North-East area, the Northwest area, the coastal area, the area of Sfax, the South East area and the South West area [17]. This affects the phonological system of the TD. In fact, the pronunciation of a word varies from one region to another. Sometimes, speakers do not pronounce certain letters while others do. For example, the city's name "منستير" "mnstyr" is often pronounced "مستير" "mstyr" by removing the letter "ن" "n". Another characteristic of the TD is the deletion of long vowels [18], especially when they are located at the end of a syllable [19]. As an example, the verb "خرج" "xarja" "he left" is pronounced "خرج" "xrj" by removing the short vowel. In addition, the consonant "ق" "q" is pronounced as "q" in the north and coastal areas, but "g" in the middle and southern areas, in most cases. But it can be pronounced "g" in the north and coastal areas for many words. For example, the city's name "قابس" "Gabes" is usually pronounced with "g" (like go) and not "qAbs." Note that Arabic examples are transliterated according to the Buckwalter Arabic transliteration.

B. Morphological Level

Morphologically, the TD has many characteristics. Some of them are inherited from MSA and others are specific to the TD. Among the morphological features specific to the TD is the use of a numeral word "زوز" "zwz" before or after the plural noun to indicate dual [20]. Another characteristic is the introduction of the proclitic "ماش" "mA$" before the verb to indicate the future. Also, this dialect includes the use of the new pronominal clitic "و" "his."

C. Syntactic Level and Word Order

Like other Arabic dialects, the TD respects in the most cases the regular grammar of MSA or even that of Classical Arabic. Bus it has some syntactic particularities which are objects of linguistic research into the syntax of dialectal Arabic that aims to find what structures are followed in such varieties [21].

Among the TD syntactic particularities, we can cite the addition of a new structure to express negation. Indeed, the TD includes the use of two particles " ما " "mA" and "ش" "$" which are located before and after the verb as in "ماخذيتش" "I do not get."

However, the TD shares some syntactic specificities of MSA like the word order flexibility of syntactic constituents. We can change the syntactic structure in a sentence and express the same meaning. Thus, starting with any word of the following example "وقتاه التران يوصل" "When does the train leaves?" does not change its meaning.

D. Orthographic Level

Since the TD is mainly spoken and not written, a word can be written in different forms depending on how it is pronounced. This variation of orthographic forms is mainly due to phonological differences in this dialect. So, the setting up of a corpus with a normalized orthographic form is a big challenge in processing Arabic Dialects.

For example, the word " آش " "|$" "what" appears in certain cases as a proclitic "ش" "$" and in certain cases it is transcribed as a separate word "آ" آش [20]. Even for the common lexical items between MSA and the TD, there are some orthographic variations. For example, the Standard Arabic word " عاصمه " "capital" can be transcribed in the TD in two ways: either by using the suffix "ة" "p" and in this case we have the same term " عاصمة " or using the suffix "ه" "h" to get " عاصمه ". All these problems are due to the absence of a standard or norm to transcribe the TD. In this context, we have proposed in previous works an OTTA (an Orthographic Transcription for Tunisian Arabic), which gathers a set of orthographic rules to manually transcribe the TD [20]. The OTTA guidelines preserve the particularities of the TD and inherit many orthographic rules from the MSA. In a recent research work, another guideline for the TD has been proposed, namely the CODA [22]. It consists of a set of orthographic rules for the TD and aims to make the TD as close as possible to MSA.

E. Lexical Level

As mentioned above, the TD is divided into sub-dialects depending on the geographical area. This classification affects the lexical level and adds enormous variation. In addition, the TD is said to be diglossic which refers to the use of MSA and the dialectal form together. It is also known to be a bilingual dialect which refers to the use of the dialect and French. So, the code switching between the Arabic and the French language affects the lexical level of the TD. In fact, it allows the addition of new dialect words which are derived from foreign languages. This explains the fact that the TD lexicon is richer than that of MSA. These are few examples: "بلاصة " " بلاس " "بقعة" to express "place." In addition to that, these foreign words are affected by the same rules to conjugate verbs in the TD. As an example, the French word "réserver" "book" is conjugated as "رزرفيلي" "rzrfyly" "book to me."

III. A Spoken Dialogue Corpus for the Tunisian Dialect

The construction of a dialogue corpus represents a big challenge especially when we deal with Arabic dialects which suffer from lack of linguistic resources [23], [24]. Even though language resources in Tunisian speech are scarce, it is possible now to find new lexical resources and new corpus developed by recent research works. A recently-developed lexical resource is the TunDiaWN, which is a new Wordnet for the TD [25]. Also, a new corpus has been recently collected is called TARIC (Tunisian Arabic Railway Interaction Corpus); it is a collection of audio recordings and transcriptions from dialogues in the Tunisian Railway Transport Network [26]. This corpus has been developed to build a phonetic dictionary in the TD. However, we can not use this corpus since it is not actually annotated in semantic concepts.

TABLE I
A Sample of a Real Dialogue in TD Between a Client (C) and a Ticket Office Clerk (S)

Turn	***Transliteration*** / **Translation**	**Transcription**
C:	*sAmHny wqtA$ yxrj ItrAn ltwns* Excuse me when does the train leave to Tunis?	سامحني وقتاش يخرج التران لتونس
S:	*mADy sAEh wrbEh OdrAj* Twenty past one	ماضي ساعه وربعه أدراج
C:	*bqdA$ hwA Itkyh* How much is the ticket?	بقداش هوا التكيه
S:	*vnA$ nlf wxmsmyh ltwns* Twelve dinars and five hundred to Tunis	ثناش نلف وخمسميه لتونس

TABLE II
Main Characteristics of the TUDICOI Corpus

# Dialogues	1825
# Speakers	1831
# Client turns	6533
# Clerk turns	5649
# Words in client turns	21682

Nevertheless, and before the development of these resources and others which are under development, we are the first who have produced, in a previous work, an initial corpus of real spoken dialogue in the TD corresponding to the task of railway inquiry [27]. This corpus is called TUDICOI, which stands for TUnisian DIalect COrpus Interlocutor. The raw version of this corpus is available online at [28]

A. Corpus Acquisition

The TUDICOI corpus was acquired with the collaboration of the National Railway Company in Tunisia (SNCFT) [29]. The recordings were made in the SNCFT main station using a digital recorder to obtain mp3 format files. This digital recorder was installed on the side of the ticket office clerk to prevent the client from knowing that the conversation was being recorded. This allowed us to avoid disruption and hesitation by clients, who then behaved in a normal way to ask for information. We had not prepared any scenario for the dialogues which were in real form and with spontaneous speech using the real form of the TD.

B. Corpus Description

The main task of the TUDICOI corpus is requesting information in the TD about the railway services. These requests are about train schedule consultation, train type, train destination, train path, fare and ticket booking. Based on these requests, several requests can be combined together during a dialogue between the ticket office clerk (S) and the client (C) about railway services in the train station. An example of a real dialogue in the TD between a client and the ticket office clerk is shown in Table I.

The TUDICOI corpus consists of 1825 dialogues from 1831 users. These dialogues represent 12182 utterances. The most important characteristics of this corpus are shown in Table II.

The 1825 dialogues are composed of 6533 client turns and 5649 clerk turns. On average, each dialogue consists of three turns for a client and three turns for a clerk. In addition, each client's turn is composed of 3.3 words, on average. It is so important to notice that the average of words per client turn is very low. On the one hand, this is due to the fact that the TD is an agglutinative language. So a word in TD could express one complete utterance. On the other hand, this is due to the exhaustive use of key words to ask for information about a limited task and omission of connectors to link words together.

In this work, we are interested only in client turns. That's why we have only labeled the transcribed utterances of clients based on a semantic point of view in order to build a language model for client utterances. In fact, we have established a well-defined semantic annotation scheme for client utterances for the railway information request task to cover all aspects of client utterances in the studied task.

C. Transcription

The transcription step appears more critical in spoken dialogue corpora because the establishment of such a resource bank represents a major effort especially when dealing with the Arabic dialect. Since there is no available speech transcription tool for Arabic [27], we have undertaken a manual transcription. In addition, it should be noted that the recording conditions were not favorable for automatic transcription. This is due to the poor quality of the acoustic signal. In fact, in addition to the conversations between the client and the ticket office clerk, there were recordings of conversations and negotiations between other staff members and clients in the train station, and there was a presence of echo caused by passengers as well as some music in the background.

The transcription was made by three transcribers. They used Arabic letters to transcribe the dialogues and they agreed on some transcription rules which consist in using the standard Arabic orthography when dealing with standard Arabic words. And when the word is specific to the TD, the word is transcribed as it is pronounced based on its phonetics. It is important to note that the transcription was made before the OTTA was proposed as a guideline to transcribe TD. For this reason, the TUDICOI corpus had to be re-processed to respect this proposed guideline.

Up to now, we have transcribed about 50% of the speech data. In fact, we recorded approximately 104 hours in the railway station and we have transcribed about 52 hours. Naturally, more than one utterance can appear in a dialogue turn. That's why we have considered a dialogue turn as a set of utterances. So, the turns were not segmented into utterances during the transcription step.

D. Corpus Processing

Generally, most of the works dealing with a speech corpus perform some automatic processing on the corpus before annotation. This processing is done to reduce the complexity of the corpus and the structures [7]. In our case, we have two versions of annotated corpus. In the first version, we did not perform any processing on the corpus before annotation. Indeed, we annotated the raw version of the client turns to look for results of discriminative models when we deal with a very raw quality of speech data. This can give important results about the robustness of discriminative models against the deteriorated data. In the second version of the annotated corpus, some automatic processing was performed before annotation to improve the turn's structure. This processing included the following tasks:

—Lexical normalization: since the transcription was done manually, any word in the TUDICOI corpus could be written in different orthographic ways. For example, we noticed that the word "رزرفسيون" "reservation" is written in four different forms: "ريزرفسيون", "رازارفسيون", "رزرفسيون", "رزرفسيون". That is why; we have performed an automatic lexical normalization which respects the OTTA guidelines presented previously.

—Morphological analysis and lemmatization: in this analysis, we have processed verbs and nouns. Verb processing is the identification of the canonical form of the verb. For example, we replace the word "خارج" "is going" and "يخرج" "goes" by the following canonical form "خرج" "go." However, noun processing consists in two steps. The first step is returning to the singular form of the noun. The second step is replacing the definite form by the undefined form of the noun. As an example, the word "الترينوات" "The trains" is transformed into "تران" "Train."

—Synonym processing: it consists in replacing each word by its synonym.

Given the lack of resources and dictionaries for the TD, we performed the lexical normalization and synonyms processing by creating a lexicon dictionary for the railway information request task. This dictionary helps us to correct the orthography and to replace each word by its synonym. Due to the absence of an automatic analyzer for the TD, we automatically performed a shallow morphological analysis for verbs and nouns by using the complete storage method [30]. For this purpose, we built a morphological base of possible changes for verbs and nouns. All these automatic treatments are not done on the same version of the annotated corpus. In fact, we prepared different versions of the annotated corpus with different levels of processing. These versions are investigated to perform multiple experiments to evaluate the CRF models on different types of processed data.

E. Annotation

In order to define a semantic scheme that covers all the concepts of the application, it is necessary to perform a semantic study about railway inquiry task to set concepts and dialogue acts. A dialogue act represents the general meaning of an utterance. A concept represents a specific minimal meaning given by a word or a group of words.

The detecting of acts was performed manually, based on the TUDICOI corpus. However, the identification of concepts was carried out in a semi-automatic way based on a statistical method. The statistical method consists in calculating the frequency of each term in the corpus. Terms which have a high frequency represent domain concepts. The only problem was with the domain terms that have a low occurrence frequency. In this case, we relied on the domain expert to specify the frequency threshold to be determined. Indeed, the domain

TABLE III
DIALOGUE ACTS USED IN THE FIRST LEVEL

Dialogue act	Example	Translation
Opening	عسلامة ، صباح الخير	Hi, good morning
Closing	بالسلامة	Bye
Undefined	شي ما عاد عندي حتى فرنك	I don't have any penny
Waiting	استنى شوية	Wait a bit
Request Information	وقتاش يخرج التران	When the train leaves
Acceptance	باهي خليهالي	Ok, leave it for me
Rejection	لالا	No No

TABLE IV
SEMANTIC CONCEPT LABELS USED IN THE SECOND LEVEL

Domain concepts		Requests concepts	
Train	Ticket_Numbers	Path_Req	Existence_Req
Train_Type	Ticket	Hour_Req	Trip_timeReq
Departure_hour	Hour_Cpt	Price_Req	Clarification_Req
Arrival_hour	Departure_Cpt		Booking_Req
Day	Arrival_Cpt	Dialogue concepts	
Origin	Price_Cpt	Rejection	Salutation_Begin
Destination	Class_Cpt	Acceptance	Salutation_End
Fare	Trip_time	Politeness	
Class	Ticket_type		
Link concepts			
Choice		Out of vocabulary	
Coordination		Out	

TABLE V
MAIN CHARACTERISTICS OF THE ANNOTATED CORPUS

# Annotated dialogues	1476
# Annotated client turns	5047
# Annotated words in client turns	16772

expert set the threshold concept and each term with a frequency value above this threshold is considered as a domain concept. To justify this choice, we conducted an empirical study which consisted in taking, each time, a small part of the corpus and calculate the frequency of threshold concept fixed by the expert. We noticed that the frequency of the threshold concept “رتور” increases with the size of the corpus but it remains the lowest value.

The result of this statistical method gives a list of domain terms. Several sets of terms represent the same concept. So, it is necessary to gather them and represent them by a single concept. For this reason, we relied on the domain expert to define for each set of terms a well defined concept.

The annotation scheme defined for concept labels respects many of the principles used in other works to annotate speech corpora, with a structure which covers the most specific details of the task. In fact, the annotation scheme used to annotate the TUDICOI corpus is inherited from the Interchange Format (IF) [31]. We have used a two-level annotation. The first level covers the general intention of the utterance by means of dialogue acts while the second level gives more specific information about the task. The second level represents the most important semantic details which are called semantic concepts.

The dialogue acts used in the first level are shown in Table III and the semantic concept labels used to label all the versions of the annotated corpus are shown in Table IV.

Given the complexity and time-consuming nature of the manual annotation task, we have only annotated 1476 dialogues. These dialogues represent 5047 client turns. The most important characteristics of the annotated corpus are shown in Table V.

The annotations were made by two annotators. For each annotated part of the TUDICOI corpus, we calculated the Kappa coefficient to guarantee an agreement between annotators. Then, annotation results are checked by a domain expert.

IV. ONTOLOGY-BASED METHOD FOR CONCEPTUAL LABELING

The lexical analysis of the TUDICOI corpus shows that utterances are characterized by the use of keywords and the absence of a definite grammatical structure. This led us to use a method guided by lexical semantics. Indeed, we try to build a knowledge base which gathers Tunisian dialect words with their semantic relations in a meaningful semantic network. The integration of a knowledge base consists in using ontologies for conceptual labeling of spoken utterances in the TD. Indeed, ontologies have been used in several systems for semantic annotation. We can cite the work of [32], which uses ontology to provide a domain-independent semantic representation of the understanding module in a dialogue system.

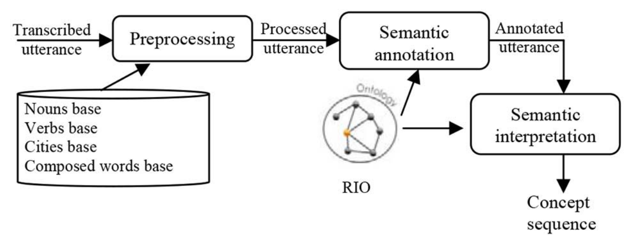

Fig. 1. Ontology-based method for conceptual labeling of TD utterances.

To build a domain ontology, we have proposed in previous work a hybrid method for semi-automatic construction of domain ontologies. This method is based on the same statistical method to extract task concepts, and a linguistic method for the identification of semantic relations between concepts. The proposed method is implemented via the ABDO tool (Assistant for Building Domain Ontology) which helps us to generate the RIO ontology (Railway Information Ontology) [33]. The RIO ontology is available online at [34].

The conceptual labeling using the RIO ontology consists of three basic steps which are: preprocessing of transcribed utterances, semantic annotation and finally semantic interpretation [35]. Fig. 1 illustrates the different steps of speech labeling based on the domain ontology.

A. Preprocessing

The main idea of this step is to treat the client’s turn in order to reduce the structure complexity and to standardize the words. It is an important step to get a correspondence between spoken words and ontology’s instances to label them by ontology concepts. This requires the use of five bases which are a noun base, a verb base, a city base, and compound word base. This processing step poses a major challenge because we are processing the TD which suffers from lack of tools and linguistic resources. So, the used bases are created manually by a domain expert during the construction phase of the RIO ontology.

اكسبراس	تونس	إلى
Train_Type	Destination Origin	Semantic relation

Fig. 2. Semantic annotation using concepts of the domain ontology.

B. Semantic Annotation

After the preprocessing step, we label utterances based on the RIO ontology concepts. In fact, we exploit our knowledge base and we seek the presence of the word which will be labeled among instances of all ontology concepts. It should be noted that it is possible that a word can be labeled by one concept, or by two concepts, or may be labeled as "O" (for Out). In the case where the word is labeled as "O", this means that the word is not present in the RIO ontology. This is due to the presence of truncated words, out of vocabulary, and other speech phenomena which we are not dealt with in this work. In the case where the word is labeled by two concepts, it is possible to improve the semantic labeling through a semantic interpretation phase presented in the following section. In the case where the word is labeled by one concept, this means that it is the correct one since this word appears as an instance for this concept in the RIO ontology.

The use of the RIO ontology for the semantic annotation step does not provide a good contribution, and it is only used as a domain dictionary. But the major contribution of the RIO integration is the exploitation of semantic relations for the interpretation step, described in the next section.

C. Semantic Interpretation

The semantic interpretation can be defined as a semantic decoding. It helps to clarify the semantic relation present in utterances to increase the accuracy of comprehension. In this step, we try to improve the previous step by exploiting the semantic relation of the RIO ontology. We should establish a correspondence between the semantic relations of the ontology and the semantic relations in the utterance. We perform the interpretation step only in the case where a word is annotated with two different concepts.

We detect all the semantic relations in the utterances during the annotation step. Then, we use this information to look for the target of the semantic relation in the RIO ontology. The target of the semantic relation is the best annotation which must be attributed to the word. To explain the proposed method for semantic interpretation, we give the following utterance as an example: "لتونس اكسبراس" "A Tunis Express." The pretreatment step transforms this utterance as follow "إلى تونس اكسبراس". Then, we perform the semantic annotation presented in Fig. 2. We notice that the word "تونس" "*Tunis*" has two different semantic concepts. It is annotated as "Destination" and as "Origin."

Now, we apply the semantic interpretation step. To well explain this further, we firstly present an extract from our ontology on a semantic network representation shown in Fig. 3. In this figure, the instance "تونس" "Tunis" belongs to two different concepts. The presence of the semantic relation "إلى" "*to*" helps to choose the correct interpretation of the word. So, the word will be annotated by the target of this relation, which is "Destination."

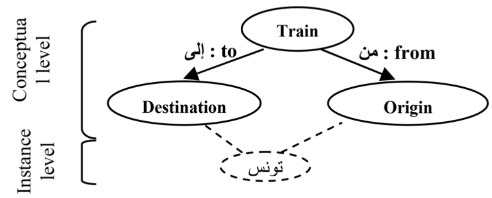

Fig. 3. Extract from the domain ontology.

During semantic interpretation to detect the semantic relations, it is possible to find many different relations. So we take a strategy to consider the closest semantic relation from the right or the left of the word to be labeled. This is due to the varied possibilities of word order in an utterance. Once a semantic relation is used in one interpretation, it is not used for other interpretations.

V. Statistical Model for Conceptual Labeling

A. Statistical Conceptual Labeling

Previous works have dealt with the conceptual labeling task by means of statistical models which have been widely used from generative to discriminative models. Raymond [4] has compared the SFST generative model to the CRF discriminative models and Wang [36] has shown that discriminative models are able to incorporate the correlated features in CRF. This advantage has reduced the error rate for spoken language understanding compared to the generative model [4]. Due to the importance of CRF models, [37] have used CRF in all stages of the understanding module, from slot filling to user intent detection. Based on their advantage, we chose the CRF models as a representative discriminative model to evaluate its performance on transcribed speech in the TD which is not segmented into utterances and with different types of processing.

Several works have dealt with the problem of conceptual labeling of the spoken language understanding using statistical methods for different languages. But these works are usually interested in dialogues which are segmented into utterances. Our proposal in these experiments is to test the performance of CRF models on the spoken TD to label the non-segmented turns. This situation seems to be more realistic since the speech recognition component provides turns as output which are not segmented into utterances. This idea is inspired from the work of Martinez [7] who performs the labeling task, using HMM in the most realistic situation where the segmentation of turns into utterances is not available.

B. CRF Based Models

CRF are undirected graphical models trained to maximize a conditional probability [38].

Lafferty *et al.* [38] define the conditional probability of a label sequence $\mathrm{y} = \mathrm{y}_1 \ldots \mathrm{y}_\mathrm{n}$ given an observation sequence $\mathrm{x} = \mathrm{x}_1 \ldots \mathrm{x}_\mathrm{n}$ as:

$$\mathrm{p}(\mathrm{y}/\mathrm{x}) = \frac{1}{\mathrm{z}(\mathrm{x})} \exp\left(\sum_{\mathrm{j}} \lambda_\mathrm{j} \mathrm{t}_\mathrm{j}\left(\mathrm{y}_{\mathrm{i}-1}, \mathrm{y}_\mathrm{i}, \mathrm{x}, \mathrm{i}\right) + \sum_{\mathrm{k}} \mu_\mathrm{k} \mathrm{s}_\mathrm{k}\left(\mathrm{y}_\mathrm{i}, \mathrm{x}, \mathrm{i}\right)\right) \tag{1}$$

With:

$$z(x) = \sum_{y} \exp\left(\sum_{j} \lambda_j t_j\left(y_{i-1}, y_i, x, i\right) + \sum_{k} \mu_k s_k\left(y_i, x, i\right)\right) \quad (2)$$

z(x) is the normalization factor that makes the sum of all probabilities equal to one. $t_j(y_{i-1}, y_i, x, i)$ represents the transition feature function of the entire observation sequence and the labels in positions i and $i-1$ in the label sequence. $s_k(y_i, x, i)$ represents the state feature function of the label in position i in the observation sequence. λ_j and μ_k are parameters which are estimated from training data. *t* and *s* are usually binary functions satisfying a certain combination of labels and observations and they are applied to each position of the sequence. These latter are defined by the user. They reflect his knowledge about the domain. They are weighted by λ_j and μ_k which give the importance of the provided information to determine the best concept sequence.

C. CRF for Conceptual Labeling of Speech in the TD

Learning CRF models consists in estimating the parameter vector $\theta = (\lambda_1, \lambda_2, \ldots \lambda_{jn}, \mu_1, \mu_2, \ldots, \mu_{kn})$ from the training data (x(i), y(i)), $i = 1 \ldots N$

Given the model as defined in Equation (1), the most probable concept sequence y^* for an input x, is:

$$y^* = \arg\ \max_y\ p(y/x) \quad (3)$$

In practice, this problem is reduced to an optimization problem. In general, it is usually solved using the quasi-Newton methods such as the L-BFGS algorithm [38].

Fig. 4 details the different steps performed to learn the CRF models for the TD from manual conceptual annotation. Then, we infer the learned models by means of a new speech which was not used in the training corpus.

After the learning step, the application of CRF to new data aims to find the most likely sequence of concepts given a new sequence of observations which has not been present in the training corpus, yet. This includes the following steps:

—Turn processing: in this step, we perform lexical normalization and synonym treatment using a lexicon dictionary for the railway inquiry task. This dictionary helps us to correct the orthography and replace each word by its synonym. Also, we perform a shallow morphological analysis for verbs and nouns.

—Automatic conceptual labeling: it is an essential step in the proposed method. It consists in labeling dialogue turns based on CRF models already learned offline. As with other stochastic methods, it is obtained with a Viterbi algorithm.

The automatic conceptual labeling step is based on CRF models. Parameters of CRF model are learned through a training corpus annotated manually. But before the annotation phase, the training corpus must be treated to normalize structure and reduce the complexity of client turns. These treatments are the same in the training step and the automatic conceptual step. Indeed, this helps us to model the same type of data on the learning level and the conceptual decoding during the test.

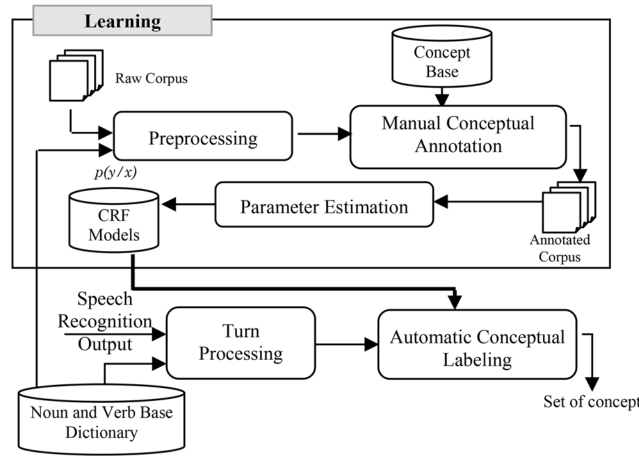

Fig. 4. CRF learning and conceptual labeling task.

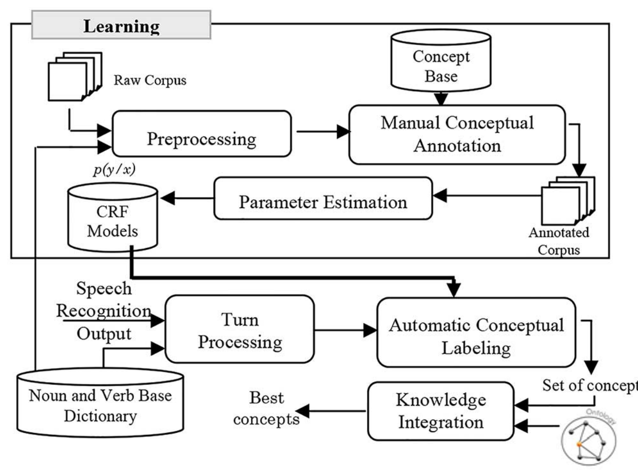

Fig. 5. Integration of the domain ontology in the labeling sequence task.

VI. Coupling Knowledge Base with a Statistical Model

Given the advantage of ontologies which gather semantic knowledge about a domain, and given the advantage of CRF models which take into account correlation and dependence between words in the same utterance, we propose a new method for robust conceptual labeling of the TD speech in the railway information task. The proposed method is based on the integration of the domain ontology with CRF models. Indeed, CRF models are very powerful in the labeling task with minimal processing on data, while the ontology provides the ability to integrate domain knowledge which is not present or just somewhat present in the training data. So the integration of a domain ontology overcomes the drawback of statistic models, and adds prior knowledge about the domain.

The proposed hybrid method for speech understanding of the TD is shown in Fig. 5. It consists of three basic steps: turn processing, automatic conceptual labeling, and integration of the domain knowledge. The first two steps are the same steps as those described above.

The knowledge integration step is the new contribution in this method. It is used to improve the annotation and semantic interpretation by the use of the RIO ontology. The domain ontology gathers specific knowledge about the railway inquiry task such

TABLE VI
RESULTS OF CONCEPTUAL LABELING BASED ON DOMAIN ONTOLOGY

CER	10.03%
Precision	84.67%
Recall	55.43%
F-measure	67%

as train schedules, cities and train stations, and other general knowledge such as days, months and numbers. This step integrates the same interpretation step based on a semantic relation of the RIO ontology, as detailed above. To cover the domain lexicon, we updated the ontology by other domain instances used in the SNCFT database.

VII. EXPERIMENTS AND RESULTS

In this section, we report the experiments that we performed with the TUDICOI corpus to test the performance of the three proposed methods (Ontology-based, CRF-based, CRF/Ontology based) described above. These experiments give us an overview about the method for conceptual labeling most suitable for under-resourced languages, since we were dealing with a small-size corpus for the TD. In our case, the conceptual labeling task consists in associating all conceptual labels to the whole dialogue turn which is not segmented into utterances. This represents the most realistic situation.

The evaluation of the obtained concepts sequence is given in terms of F-measure and Concept Error Rate (CER). The CER is computed as an incorrect prediction with reference concepts, and the F-measure is a combination between precision and recall.

$$\mathrm{F-measure} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \times 100 \quad (4)$$

$$\mathrm{CER} = \frac{\#\mathrm{incorrect\_\ concept\_pr\ ediction}}{\#\mathrm{reference\_\ concepts}} \quad (5)$$

A. Results of the Ontology–based Conceptual Labeling

In this section, we evaluate the performance of the conceptual labeling method based only on the RIO ontology.

To populate the ontology with terms used in railway inquiry, we used a training corpus which consists of 13555 words. Then, we used a test corpus composed of 3217 words to evaluate the proposed method based only on the domain ontology. The results of the conceptual labeling task based on the RIO ontology are reported in Table VI.

From the obtained results, we can conclude that the ontology-based method is able to perform automatic conceptual labeling. Indeed, the RIO ontology helps us to find the correct conceptual label for the words through the exploitation of the ontology's concepts as it helps to link them through semantic relations detected at the same time in the turn and in the ontology.

However, the failure of this method is evident in the case of the absence of lexical semantic relations in the turn, which complicates the conceptual labeling task. This is due to the fact that the client's utterances are generally based on intonation to indicate their requests (a question, a clarification, an acceptance, etc). An instance of the failure of the ontology-based method is presented in the following example: "تونس برميير" "*Tunis First Class*." In this example, the word "تونس" "*Tunis*" is labeled by two different concepts because of the absence of a relationship to indicate if it is a destination or an origin. But the most probable case is to consider this word as a destination city. That's why we integrate the use of stochastic knowledge to model the most likely case based on a training corpus.

TABLE VII
TUDICOI CORPUS SIZE

	Training	Test
Dialogues	1202	267
Turns	4131	906
Words	13555	3217

B. Results of CRF-based Conceptual Labeling

There are many works which dealt with the conceptual labeling task using statistical models. The majority of these works used segmented turns with big data [39], and few of them used unsegmented turns [7]. In our knowledge, there are no works that handled CRF models with unsegmented dialogue turns. This is the most realistic case since a speaker's turn is the output of the recognition module in a dialogue system. In addition, dealing with unsegmented turns reduces the manual segmentation of turns into utterances. For this reason, we would like to test the performance of CRF models on unsegmented dialogue turns in the TD learned from a small-size corpus.

For these experiments, we have prepared three versions of the annotated corpus. The first one is a raw corpus (called Set I) which is not been treated beforehand. The raw version of the TUDICOI corpus presents many complex problems. In fact, words do not respect the same orthography, so, we can find the same word transcribed with different forms. Also, we can find many morphological problems due to the TD features. As an example, a word can be agglutinated with other words. Experiments on the raw version of the TUDICOI corpus, helps us to test the performance of CRF models on complex and unsegmented turns. In the second version of the annotated corpus (called Set II), we performed a morphological analysis for the verbs and the nouns, a lexical standardization and a synonym processing. In the third version of the annotated corpus (called Set III), we ameliorated the second version. In fact, we performed a morphological analysis for the verbs and the nouns, a lexical standardization and a synonym processing. Then, we treated the cities' names. This consists in dissociating the concept marker from the city's name and gathering a city's name (as a compound word) if it is composed of two words. As an example of compound city's name is "بير بورقبة" "Bir Bouregba."

We used the exhaustive cross validation to divide the corpus into two parts. The first part of the annotated TUDICOI corpus represents 80% which is used for the training and the remaining 20% is used for the test. Table VII shows the characteristics of these parts.

After the manual annotation step, we converted the annotated corpus into a standard annotation adopted by CRF. In fact, we had to convert the conceptual annotation into sequence labels in order to represent this as a sequence labeling problem. The

TABLE VIII
CONCEPTUAL LABELING RESULTS OF SET I (RAW DATA)

	Template0	Template1	Template2	Template3
CER (%)	20.05	10.44	8.98	8.47
Precision (%)	79.19	88.50	89.86	90.33
Recall (%)	76.36	80.45	79.68	79.21
F-measure (%)	77.75	84.28	84.46	84.40

TABLE IX
CONCEPTUAL LABELING RESULTS OF SET II

	Template0	Template1	Template2	Template3
CER (%)	18.65	9.05	8.39	7.70
Precision (%)	80.72	90.22	90.75	91.45
Recall (%)	78.13	83.57	82.47	82.47
F-measure (%)	79.40	86.77	86.41	86.73

TABLE X
CONCEPTUAL LABELING RESULTS OF SET III

	Template0	Template1	Template2	Template3
CER (%)	20.05	9.19	8.51	7.86
Precision (%)	79.28	90.12	90.66	91.26
Recall (%)	76.75	83.93	82.68	82.24
F-measure (%)	77.99	86.92	86.48	86.52

standard scheme adopted in CRF respects the fact that multiple words can be annotated with the same conceptual label. This label scheme is called BIO annotation. The label starting with (B-???) refers to the beginning of the conceptual segment, (I-???) refers to any word in the conceptual segment, and finally, O is for words that do not refer to any conceptual label. This annotation is able to segment two different conceptual labels of the same type appearing side-by-side [5].

In the first experiment, we used CRF models for conceptual labeling, i.e. associating a conceptual label to each group of words in the turn. Then, we ameliorated the features introduced to CRF models. In fact, we began our experiments with a simple template (Template 0) which represents a simple semantic annotation. It is only based on states and transitions of CRF models only. Then, we improved the first result by adding unigram features which take into account the previous and the next word in the turn (Template 1). Then, we added to Template 1 another feature which incorporates two words before and after the current word to be labeled (Template 2). Finally, we incorporated in Template 2 correlated features which take into account the correspondence between the current word and the previous one and the current word and the next one (Template 3).

After the preparation of the different versions of the annotated corpus and the different types of templates, we learned the CRF models to check their performance in the different types of treated data in the TD. The results are reported in Tables VIII, IX and X.

TABLE XI
CHARACTERISTICS OF THE TEST CORPUS

#Dialogues	#User words	Client turns			
		A	D	X	Total
267	3217	379 41.83%	482 53.21%	45 4.96%	906

TABLE XII
CER RESULTS USING TEMPLATE3 ACROSS THE DIFFERENT SETS A, D AND X AND WITH DIFFERENT TYPES OF PROCESSED DATA (SET I, SET II AND SET III)

	Set I	Set II	Set III
A	7.29%	6.15%	5.74%
D	9.92%	9.66%	10.61%
X	10.52%	8.78%	8.77%

In our experiments, we used the CRF ++ toolkit [40]. It is a simple, customizable and open source implementation of CRF for segmentation and labeling of sequence data.

Tables VIII, IX and X compare the conceptual error rate (CER) and the F-measure with different templates set across different types of data.

For all types of data (Set I, Set II and Set III), the CER significantly decreases across the different types of templates. So, the amelioration of templates affects the number of features to train CRF models. However, the differences of CER across different types of preprocessed data are not very significant. This proves that CRF models perform well with slightly treated data. In fact, the CER does not decrease significantly between different types of data especially between Set II and Set III. This is a very important result since CRF models perform well with minimal preprocessing on the training data. This shows the robustness of such a model with noisy data in comparison with a knowledge-base method which requires robust preprocessing for the training data. In addition, and through a manual examination of automatic labeling result using CRF, we found that the CRF models have the ability to detect the composed tokens specific to the task and label them correctly.

We have also performed other experiments based on the CRF models learnt above. We classified the set of client turns into three types of requests, according to the standards proposed by the ARPA community [41]. These sets are context-independent requests (Set A), context-dependent requests (Set D) and finally out of context requests (Set X).

Table XI gives an overview of the different sets. These experiments helped us to know the source of the error. Obtained results are shown in Table XII. It is clear that errors are due to the presence of out of context utterances (Set X) and context dependent utterances (Set D), since we are interested in the literal conceptual labeling and not the contextual one. That's why context dependent utterances increase the CER.

As a conclusion, and despite the raw quality of the first set, CRF performs well in comparison with the automatic conceptual task using a knowledge base method. In fact, CRF do well compared to the results reported in Table VI using the domain ontology as knowledgebase with well-performed data prepared in advance.

TABLE XIII
LABELING RESULTS USING CRF MODELS (USING TEMPLATE3 ON SET III) WITH DOMAIN ONTOLOGY INTEGRATION

Precision	Recall	CER	F-measure
92.27%	84.94%	7.11%	88.45%

TABLE XIV
COMPARISON BETWEEN OUR PROPOSED METHODS

Method	F-measure	CER
RIO Based	67.00%	10.03%
CRF Based	86.52%	7.86%
CRF/RIO Based	88.45%	7.11%

Through a manual examination of automatic conceptual labeling data using CRF models, we have detected another source of failure. This failure is due to the presence of new words which belong to the lexicon of the task, and which are not presented in the training corpus. This is already the major drawback of the statistical models. As an example, the city's name "جرجيس" "Jarjiss" is annotated by CRF models as "O" (Out) because it does not appear in the training corpus. Other examples can be cited such as the day's name, the month's name, numbers and times which can be found in the SNCFT database. This is due to the small size of the corpus used to learn the model. So, we integrated a domain ontology to solve this problem.

C. Statistical Results with Integration of a Domain ontology

In this section, we evaluate the proposed hybrid method based on the coupling between CRF models and the domain ontology.

In these experiments, we used the same CRF models already learned using the CRF++ tool. Also, we used the RIO ontology generated by the ABDO tool (Assistant for Building Domain Ontology), and we updated it with other bases from the SNCFT company. To learn CRF models, we used the same training corpus used in the experiments above. Indeed, this corpus consists of 13555 words annotated in terms of domain concepts for training, and we used 3217 words as test corpus. In addition, the turns used for training and for the test are not segmented into utterances. Table XIII reports the conceptual labeling evaluation based on coupling CRF models with the domain ontology.

The hybrid method proposed for the conceptual labeling of turns in the TD performs well by decreasing the error rate and increasing the F-measure in comparison with a method based only on CRF models (from 7.86% to 7.11% as CER and from 86.52% to 88.45% as F-measure). Indeed, the integration of a knowledge base through the ontology has improved the performance of CRF models by adding new words which does not appear in the training corpus. These new words represent about 3% of the test corpus. Indeed, the integration of the ontology has reduced the number of incorrect predictions, from 219 to 198, which helps to reduce the CER of 0.75%.

In order to compare the different methods proposed, we present in Table XIV a summary of results obtained.

In our method, we perform the understanding task in the most realistic form where dialogue turns are not segmented into utterances. In addition, we model CRF from a small-size corpus. The RIO based method does not perform well for the conceptual labeling task compared to the CRF or CRF/RIO based methods. In addition, building the RIO ontology is not an easier task than building a manually annotated corpus. This is due to the tedious task to define semantic relations by a linguistic expert. However, the ontology annotation is easier than manual conceptual labeling to obtain the training corpus. The RIO based method has failed in the case where semantic relations have not been present in the utterance. That's why the two other methods perform well in comparison with the RIO based method.

Finally, we can conclude that the proposed hybrid method for speech understanding is efficient in the case of the TD which suffers from lack of linguistic resources and tools for automatic processing. In fact, we have used a small-size corpus to learn CRF models. Thereafter, we have improved CRF performance by the integration of the domain ontology which allows adding prior knowledge about the task. Indeed, the CRF models can not take into account the words rarely used in the training corpus, as they do not recognize the new vocabulary. The integration of the ontology has overcome the failure of the CRF models by adding a knowledge layer about the task.

VIII. CONCLUSION

In this paper, we proposed an original method for speech understanding of the TD. This method is based on the coupling between the CRF models and domain ontology. CRF models perform conceptual labeling. These models take into account the dependency between the words in a probabilistic model. However, the domain ontology gathers all the lexicon used in the task, even those which do not appear in the training corpus. Our method is applied to the most realistic case where turns are not segmented into utterances. The obtained results are interesting. Indeed, we have increased the F-measure and decreased the error rate in comparison with the RIO-based method or CRF-based method. The insufficiency of CRF models is due to the small size of the corpus. So the integration of the domain ontology helps to ameliorate the conceptual labeling task, by adding a new lexicon of the domain. This idea is very interesting not only for Arabic dialects which suffer from lack of linguistic resources but also for other under-resourced languages.

REFERENCES

[1] H. Aust, M. Oerder, F. Seide, and V. Steinbiss, "The Philips automatic train timetable information system," *Speech Commun.*, vol. 17, no. 3–4, pp. 249–262, Nov. 1995.

[2] J. Dowding, J. M. Gawron, D. Appelt, J. Bear, L. Cherny, R. Moore, and D. Moran, "Gemini: A natural language system for spoken language understanding," in *Proc. ARPA Workshop Human Lang. Technol.*, Princeton, NJ, USA, 1993.

[3] W. Ward and S. Issar, "Recent improvements in the CMU spoken language understanding system," in *Proc. ARPA HLT Workshop*, 1994.

[4] C. Raymond and G. Riccardi, "Generative and discriminative algorithms for spoken language understanding," in *Proc. Interspeech*, Antwerp, Belgium, 2007.

[5] C. Sutton and A. McCallum*, An Introduction to Conditional Random Fields for Relational Learning*. Cambridge, MA, USA: MIT Press, 2006.

[6] S. Furui, "Recent progress in spontaneous speech recognition and understanding," in *Proc. IEEE Workshop Multimedia Signal Process.*, 2002, pp. 253–258.

[7] C. D. Martínez-Hinarejos, B. José-Miguel, and G. Ramón, "Statistical framework for a Spanish spoken dialogue corpus," *Speech Commun.*, vol. 50, pp. 992–1008, 2008.

[8] B. José-Miguel, L. Eduardo, V. Amparo, M. Castro, I. Galiano, R. Justo, I. Lopez de Letona, and A. Miguel, "Design and acquisition of a telephone spontaneous speech dialogue corpus in Spanish: Dihana," in *Proc. LREC*, Genoa, Italy, 2006, pp. 1636–1639.

[9] M. Dinarelli, A. Moschitti, and G. Riccardi, "Discriminative reranking for spoken language understanding," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 526–539, Feb. 2012.

[10] F. Sha and F. Pereira, "Shallow parsing with conditional random fields," in *Proc. HLT-NAACL*, 2003, pp. 134–141.

[11] Y. Y. Wang and A. Acero, "Discriminative models for spoken language understanding," in *Proc. ICSLP*, 2006.

[12] F. Duvert and R. D. Mori, "A conditional model for triggering understanding actions in a speech understanding system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2011, pp. 5620–5623.

[13] A. Zouaghi, M. Zrigui, and M. Ben Ahmed, "Évaluation des performances d'un modèle de langage stochastique pour la compréhension de la parole arabe spontanée," in *Proc. TALN*, 2007.

[14] Y. Bahou, L. Hadrich Belguith, and A. Ben Hamadou, "Towards a human-machine spoken dialogue in Arabic," in *Proc. LREC*, 2008.

[15] M. Diab and N. Habash, "Arabic dialect processing tutorial," in *Proc. Human Lang. Technol. Conf. North Amer.*, Rochester, NY, USA, 2007.

[16] A. Khalfaoui, "A cognitive approach to analyzing demonstratives in Tunisian Arabic," Ph.D. dissertation, Univ. of Minnesota, Minneapolis, MN, USA, 2009.

[17] B. Ouerhani, "Interférence entre le dialectal et le littéral en Tunisie: Le cas de la morphologie verbale," *Synergies*, vol. 1, pp. 75–84, 2009.

[18] M. Tilmatine, "Substrat et convergences: Le berbère et L'arabe nord-africain," *Estudios de dialectologia norteafricana y andalusl*, vol. 4, pp. 99–119, 1999.

[19] S. Mejri, M. Said, and I. Sfar, "Pluringuisme et diglossie en Tunisie," *Synergies*, vol. 1, pp. 53–74, 2009.

[20] I. Zribi, M. Graja, M. Ellouze Khmekhem, M. Jaoua, and L. Hadrich Belguith, "Orthographic transcription for spoken Tunisian Arabic," in *Proc. CICLing'13*, 2013, pp. 153–163.

[21] K. E. Brustad, *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, DC, USA: Georgetown Univ. Press, 2000.

[22] I. Zribi, R. Boujelbane, A. Masmoudi, M. Ellouze, L. Hadrich Belguith, and N. Habash, "A conventional orthography for Tunisian Arabic," in *Proc. LREC*, Reykjavik, Iceland, 2014, pp. 2355–2361.

[23] N. Habash, M. Diab, and O. Rambow, "Conventional orthography for dialectal arabic," in *Proc. LREC*, Istanbul, Turkey, 2012, pp. 711–718.

[24] R. Zbib, E. Malchiodi, J. Devlin, D. Stallard, S. Matsoukas, R. M. Schwartz, J. Makhoul, O. Zaidan, and C. Callison-Burch, "Machine translation of Arabic dialects," in *Proc. HLT-NAACL*, 2012, pp. 49–59.

[25] R. Bouchlaghem, A. Elkhlifi, and R. Faiz, "Tunisian dialect Wordnet: Creation and enrichment using web resources and other Wordnets," in *Proc. EMNLP Workshop ANLP*, Doha, Qatar, 2014, pp. 104–113.

[26] A. Masmoudi, M. Ellouze Khmekhem, Y. Esteve, L. Hadrich Belguith, and N. Habash, "A corpus and phonetic dictionary for Tunisian Arabic speech recognition," in *Proc. LREC*, Reykjavik, Iceland, 2014, pp. 306–310.

[27] M. Graja, M. Jaoua, and L. Hadrich Belguith, "Discriminative framework for spoken Tunisian dialect understanding," in *Lecture Notes in Computer Science LNCS*. New York, NY, USA: Springer, 2013, vol. 7978, pp. 102–110.

[28] The raw version of the TuDiCoI corpus, [Online]. Available: https://sites.google.com/site/marwagraja/resources

[29] [Online]. Available: www.sncft.com.tn/

[30] V. Clavier and G. Lallich-Boidin, "Modélisation linguistique de la suffixation en vue de l'analyse automatique," *Traitement Automatique des Langues*, vol. 35, no. 2, pp. 129–144, 1994.

[31] N. Alcacer, J. Benedi, F. Blat, R. Granell, C. D. Martinez, and F. Torres, "Acquisition and labeling of a spontaneous speech dialogue corpus," in *Proc. SPECOM*, Patras, Greece, 2005, pp. 583–586.

[32] J. Allen, M. Dzikovska, M. Manshadi, and M. Swift, "Deep linguistic processing for spoken dialogue system," in *Proc. Workshop Deep Linguist, Process. (DeepLP'07)*, 2007, pp. 49–56.

[33] J. Karoui, M. Graja, M. Boudabous, and L. Hadrich Belguith, "Domain ontology construction from a Tunisian spoken dialogue corpus," in *Proc. ICWIT*, Hammamet, Tunisia, 2013.

[34] The RIO Ontology, [Online]. Available: https://sites.google.com/site/marwagraja/resources

[35] M. Graja, M. Jaoua, and L. Hadrich Belguith, "Towards understanding spoken Tunisian dialect," in *Lecture Notes in Computer Science*. New York, NY, USA: Springer, 2011, pp. 131–138.

[36] T. Wang, J. Li, Q. Diao, Y. Z. Wei Hu, and C. Dulong, "Semantic event detection using conditional random fields," in *Proc. IEEE CVPRW'06*, 2006.

[37] A. Deoras, G. Tur, R. Sarikaya, and D. Hakkani-Tür, "Joint discriminative decoding of words and semantic tags for spoken language understanding," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 8, pp. 1612–1621, Aug. 2013.

[38] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2001, pp. 282–289.

[39] A. Stolcke, N. Coccaro, R. Bates, P. Taylor, P. C. van Ess-Dykema, K. Ries, E. Shriberg, D. Jurafsky, R. Martin, and M. Meteer, "Dialogue act modeling for automatic tagging and recognition of conversational speech," *Comput. Linguist.*, vol. 26, no. 3, pp. 1–34, 2000.

[40] T. Kudo, *CRF++*, 2005 [Online]. Available: http://chasen.org/~taku/software/CRF++/

[41] W. Minker and S. Bennacef, *Speech and Human Machine Dialog*. Boston, MA, USA: Kluwer, 2004.

M. Graja received her Engineer degree in computer science engineering and the Master degree from the National School of Engineers of Sfax-Tunisia (ENIS), respectively, in 2006 and 2008. Currently, she is pursuing her Ph.D. studies with the Arabic Natural Language Processing Research Group (ANLP-RG) of MIRACL Laboratory. Her research interests include Tunisian dialect and dialogue systems.

M. Jaoua received his Ph.D. degree in computer science from the University of Tunis in 2004. He is currently an Assistant Professor of computer science at the Faculty of Economic Sciences and Management of Sfax (FSEGS)-University of Sfax (Tunisia). He has been a member of ANLP-RG within MIRACL Laboratory since 1997. His main research activities are text summarization and dialogue systems.

L. Hadrich Belguith received her Ph.D. degree from the Faculty of Sciences -Tunis, Tunisia, in 1999. Currently, she is a Professor of Computer Science and Management at FSEGS and the head of the ANLP-RG of MIRACL laboratory. Her research activities have been devoted to several topics: Arabic text analysis, Automatic Abstracting, Question-Answering systems, and Human–machine spoken dialogue systems.