Deep Study of CRF Models for Speech understanding in Limited Task

Marwa Graja

College of Computer and Information Sciences
Jouf University, KSA

Abstract—**In this paper, we propose to evaluate in depth CRF models (Conditional Random Fields) for speech-understanding in limited task. To evaluate these models, we design several models that differ according to the level of integration of local dependencies in the same turn. As we propose to evaluate these models on different types of processed data. We perform our study on a corpus where turns are not segmented into utterances. In fact, we propose to use the whole turn as one unit during training and testing of CRF models. This represents the natural way of conversation. The language used in this work is the Tunisian Arabic dialect. The obtained results prove the robustness of CRF models when dealing with raw data. They are able to detect the semantic dependency between words in the same speech turn. Results are important when CRF models are designed to take into account the words with deep dependencies in the same turn and with advanced preprocessed data.**

Keywords—Speech understanding; Arabic dialect; CRF models

I. INTRODUCTION AND RELATED WORKS

Spoken Language Understanding is an important component in spoken dialogue systems. It aims to extract concepts from an utterance to clarify speech meaning. Therefore, the key link in an automatic understanding process revolves around the correspondence between the set of words in the utterance and the set of semantic concepts. In order to resolve this correspondence, the first research works in this field exploited linguistic formalisms such as regular grammars and context-free grammars. Recent works have rather oriented towards the exploitation of machine learning models for concept detection, these models are widely used for the semantic annotation of speech utterances.

Our overview of the literature showed that learning models constitute the dominant context for speech understanding due to the performances recorded particularly in restricted domains. These models enjoy several advantages reported by [1] [2] [3]. Indeed, the intervention of a human expert is limited to the labeling of data, which represents an easier task than the modeling of grammars or patterns. Moreover, these models offer better portability since they are domain and language independent. However, the effectiveness of these models is sensitive to the used corpus, which must be representative and large, in order to determine their parameters [4].

Machine learning models are classified into generative and discriminative models [2] and they are widely applied to speech understanding. HMM (Hidden Markov Models) is an example of generative models and they are used by [5] for speech understanding of Spanish language, using the DIHANA corpus. The DIHANA corpus task deals with requests of information about railway services. This work uses HMM in the most realistic situations where dialogues are not segmented into utterances. The results of their work are very important. They obtained 92% as F-measure. This good result is due to the large size of the corpus used for training models.

In the literature, several studies show that discriminant models perform better than generative [6] [2] [7]. CRF models (Conditional Random Fields), as an example of discriminant models, have been widely exploited in many tasks in natural language processing such as semantic annotation and syntactic analysis [8], [9] [10]. A particular distinction is reported for CRF models whose performance exceeded that of other models, [11] [2] [12]. It is so important to notice that the CRF models have the capacity to integrate correlated characteristics that make it possible to take into account the local context of an utterance. All of these observations encouraged us to exploit these models in the context of the speech understanding of the Arabic dialect.

Several works have shown the robustness of Conditional Random Fields (CRF) models for request information in the French language using the MEDIA corpus [12]. The MEDIA corpus is manually annotated with semantic concepts of touristic information. Turns in this corpus are segmented into utterances, which simplifies the speech understanding. Raymond et al. [12] have used CRF models and 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.

In addition, CRF models offer two major advantages. On the one hand, they allow segmentation and conceptual annotation taking into account the local context of the utterance. On the other hand, they make it possible to guarantee convergence towards the most probable concepts by taking into account all the previous and following observations in the statement [13]. Indeed, these models have the ability to use all the observations of a sequence to predict a conceptual label. This represents an interesting distinction compared to HMM (Hidden Markov Models).

In this paper, we propose to evaluate the performance of CRF models. We designed several models that differ according to the level of integration of local dependencies in the same turn. We also propose to use several processing levels on the corpus. In addition, almost all learning-based understanding

methods have been interested in modeling speech turns segmented into utterances, we suggest to use turn as a whole unit (not segmented into utterances) to test the performance of CRF models, which represents the natural way of conversation.

This paper is organized as follows. In Section II, we present CRF models for speech understanding. Section III presents Spoken Dialogue Corpus for Tunisian Dialect. Section IV deals with evaluation metrics. In Section V, we present experiments and discussion. A conclusion is drawn in Section VI.

II. CRF MODELS FOR SPEECH ANNOTATION

Conditional Random Fields (CRFs), initiated by Lafferty, are discriminant models that define the conditional probability of observation sequences according to label sequences [14]. Lafferty defines the conditional sequence labeling probability Y=y1…yn given an observation sequence X=x1…xn as follows:

$$P(Y|X) = \frac{1}{z(X)} \exp(\sum_j \lambda_j t_j(y_{i-1}, y_i, X, i) + \sum_k \mu_k s_k(y_i, X, i)) \quad (1)$$

With

$$z(X) = \sum_y \exp(\sum_j \lambda_j t_j(y_{i-1}, y_i, X, i) + \sum_k \mu_k s_k(y_i, X, i)) \quad (2)$$

- z(X) is a factor that normalizes the probabilities.

- t_j (y_{i-1}, y_i, X, i) represents the transition characteristic function for an observation sequence between the labels at position i and i-1.

- s_k (y_i, X, i) represents the characteristic function of the state of the label for a sequence of observations .

- λ_j et μ_k are real values which make it possible to attribute a weight to each characteristic function to specify its importance. These values therefore make it possible to characterize the discriminating power of the model. These parameters are fixed during the learning phase and make it possible to maximize the likelihood on a set of already annotated data.

Referring to the model defined in Eq. (1), the most likely sequence of concepts for labeling a sequence of input words is:

$$Y^* = \arg\max_y P(Y|X) \quad (3)$$

CRF is modeled by undirected graph models (see Fig. 1) to define a probability distribution over a process of label Y given an observation X, by maximizing a conditional probability [14].
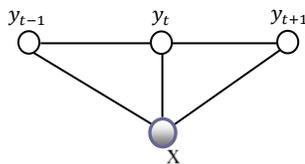
Fig. 1. Graphical design for CRF models

In this graph, the set of vertical nodes are two random fields X and Y, respectively describing the set of observations and the set of annotations. Two variables linked in the graph express that one depends on the other. Based on this, each node y_t depends on the preceding node y_{t-1} and the following node y_{t+1}, and implicitly on the variable x. Therefore, each variable y_t must be linked to a variable x to guarantee the dependence between the labels on the one hand and the sequences of observations on the other.

Learning CRF models consists in determining from the learning corpus, the vector $\theta = \{\lambda_1, \lambda_1, \dots, \lambda_{k1}, \mu_1, \mu, \dots, \mu_{k2}\}$ that represents the weight vector of the characteristic functions t and s. After the learning step, the exploitation of CRF models on new data consists in finding the most probable sequence of states given a new sequence of observations, which are not encountered in the training corpus. We perform this by applying the Viterbi algorithm, as it is the case with HMM models.

III. SPOKEN DIALOGUE CORPUS FOR TUNISIAN DIALECT – TUDICOI

A. Corpus Description

The TUDICOI corpus, used in this work, consists of spontaneous oral dialogues of railway request information, in Tunisian Dialect (TD). The purpose of these requests is to consult the timetables of the train, the type of train, the destination of the train, the route taken by the train, the price and types of tickets and the reservation of tickets. We should notify that several requests can be combined during a dialogue between tellers and customers [15].

The transcribed part of the TuDiCoI corpus consists of 1825 dialogues representing 12182 turns. These turns consists of 6533 customer turns and 5649 agent turns. We list the features of the TuDiCoI corpus in Table I.

TABLE I. SPECIFICATION OF THE TUDiCoI CORPUS

# dialogues	1 825
# customer turns	6 533
# agent turns	5 649
# customer word	21 551
# customer vocabulary	1 437

Note that on average, each dialogue consists of three turns for the customer and three turns for the agent. Additionally, each customer turn is comprised of an average of 3.3 words. It is important to note that this average is low due to the agglutinative aspect of words in the dialect and the frequent use of keywords to request for information.

TABLE II. CHARACTERISTICS OF SOME DIALOGUE CORPORA IN LIMITED TASKS

Corpus	#D	#T	#V	#W	A	L	Task
Trains 93	98	5 900	860	55 000	H-H	Ang	Manufacture and shipment of goods on the railways.
DIHANA	900	15 413	823	48 243	H-M	Esp	Railway information.
SARF	350	9 763	827	117 156	H-M	Ar	Railway information.

TuDiCoI	1 825	12 182	1 437	21 551	H-H	TD	Railway information.
TARIC	4 662	18 657	--	71 684	H-H	TD	Railway information.
MEDIA	1 257	38 434	2 715	156 048	H-M	Fr	Hotels reservation.

Table II shows the characteristics of different dialogue corpora used in other projects in different languages. It should be noted that #D designates the number of dialogues, #T the number of speaking turns, #V the size of the vocabulary, #W the number of words. A designates the type of corpus (H-H for Human-Human and H-M for Human-Machine). Finally, L provides information on the language used (Eng. for English, Esp. for Spanish, Ar. for Arabic, TD for the Tunisian dialect and Fr. for French). These corpora vary in size from a few tens to thousands of dialogues.

B. Annotation Schema

We proposed in this work an annotation scheme to perform the manual concept annotation step. Table III summarizes the annotation scheme defined for dialogue acts and semantic concepts [16].

TABLE III.   SEMANTIC CONCEPTS

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		
Destination	Class_Cpt	Rejection	Salutation_Begin
Fare	Trip_time	Acceptance	Salutation_End
Class	Ticket_type	Politeness	
Link concepts			
Choice		Out of vocabulary	
Coordination			
		Out	

Due to the complexity of the annotation task, effort, and manual verification, we have annotated only 1476 dialogues which represents 5047 customer turns. The characteristics of the annotated corpus are summarized in Table IV.

TABLE IV.   CHARACTERISTICS OF THE ANNOTATED PART OF THE TUDICOI CORPUS

# Annotated dialogues	1 476 / 1 825
# Customer annotated turns	5 047 / 6 533
# Annotated Customer words	16 772 / 21 551

In order to define the parts of the TuDiCoI corpus used for our evaluations, we have divided the annotated corpus into two parts. The first part of the corpus is used for learning and it represents about 80% of the total size, while the second part constitutes 20% of the corpus used for the test. Table V provides information on the characteristics of these two different parts in terms of number of dialogues, speaking turns and words.

TABLE V.   CHARACTERISTICS OF TRAINING AND TEST CORPUS

	Training	Test
Dialogues	1202	267
Turns	4131	906
Words	13555	3217

Since we are interested in literal understanding, which does not depend on dialogical context, we have classified all the speaking turns of the test part into three types, according to the recommendation proposed by the ARPA community [17], namely sets A, D and X. Table VI presents the characteristics of these different sets. This classification makes it possible to give an overview of the types of turns contained in the test part.

TABLE VI.   CHARACTERISTICS OF THE TEST CORPUS DIVIDED INTO THREE SETS A, D AND X

# Client turns			
A	D	X	Total
379	482	45	906
41.83%	53.21%	4.96%	100%

The first set corresponds to the context-independent customer turns (Set A). This set contains the turns that have no connection with the history of the dialogue. While the second set corresponds to the context-dependent ones (Set D). This set contains the turns that have a relationship with the dialogical context. The third set corresponds to out-of-context turns of the dialogue (Set X). This set includes marginal turns that are not related to the domain. Table VII shows an example for each series.

TABLE VII.   EXAMPLE FOR EACH SET A, D AND X

Set	Transcription ◀ / ▶Translitteration / Translation
A	تكيه لتونس برميير كلاس / tikiyh ltuwnis brumyiyr klaAs / *A first class ticket to Tunis*
D	بقداش هي ؟ / b.qad~Aš hiya ? / *How much does it cost ?*
X	دخلت عالأنترنات وشفت / dxalt ςalÂantirnaAt wušuft / *I am connected to the Internet and I saw*

We utilize these different sets of the TuDiCoI corpus for the evaluation of CRF-based speech understanding method for the TD.

Almost all speech-understanding methods are interested in modeling speech turns segmented into utterances. The alternative we have proposed is to use the turn as a whole unit for training and testing the performance of CRF models [16]. This represents the natural way of conversation.

C. Pretreatments

To evaluate CRF-based speech understanding, we prepared three versions of the TuDiCoI corpus:

- The first version (version I) is a raw version which is not pre-processed, thus increasing the complexity of the structure of the dialect turns. In this version, the words do not respect the spelling transcription guide. Therefore, a word can be written in different spellings. Likewise, this version has morphological problems

such as the agglutination of a particle with the word, which follows it. The evaluations carried out on this version of the corpus is used to test the performance of the CRF models on data not processed in advance.

- The second version of the annotated corpus (version II), is partially preprocessed. This version has undergone spelling correction, morphological analysis of verbs and nouns, as well as synonymy analysis processing.

- The third version of the annotated corpus (version III) presents an improvement compared to the second version and which consists in processing the agglutinations of the names of cities, which makes it possible to dissociate the particle, if it exists, from the name that is attached to it.

D. Tabular Corpus

After the manual labeling step into concepts, we converted each version of the annotated corpus into a standard representation adopted by CRF models. This representation uses a set of labels called the BIO notation (Begin Inside Outside) [18], in which:

- The label starting with "B-???" indicates the beginning of the conceptual segment.

- The label "I-???" denotes any meaningful word that is part of the conceptual segment.

- The label "O" is assigned for words that do not refer to any conceptual label.

سلام	B- Salutation_Begin
عليكم	I- Salutation_Begin
تكيه	B-Ticket
لتونس	B-Destination
بقداه	B-Price_Req
زوز	B-Ticket_Numbers
زوز	B-Ticket_Numbers
لتونس	B-Destination
وقتاش	B-Dep_Hour_Req
يخرج	B-Departure_Cpt
هو	O
تسكره	B-Ticket
لتونس	B-Destination
ماضي	B-Dep_hour
ساعه	I-Dep_hour

Fig. 2. Example of BIO notation from the TUDICOI corpus

The advantage of using BIO notation is that it is able to segment a set of words into several conceptual segments and display them one after the other [19] [7]. An example from TUDICOI corpus in BIO notation is shown in Fig. 2.

The tabular corpus is used for training and testing CRF models. Conceptual labeling using CRF models consists in finding the best sequence of states, given a sequence of input observations. This problem is solved using the Viterbi algorithm due to the linear topology of CRF models [20]. This algorithm makes it possible to give the list of n best results.

IV. EVALUATION METRICS

The evaluation makes it possible to evaluate the conceptual correspondence, which consists in seeking the pairing between a set of words of a turn and a set of semantic concepts. For this, we use the Concept Error Rate (CER). The CER makes it possible to compare the list of reference semantic concepts with the list of concepts emitted by the system according to the following equation:

$$CER = \frac{\#\ incorrect\ predicted\ concepts}{\#\ reference\ concepts} \qquad (4)$$

We use other measures to evaluate conceptual labelling such as Precision, Recall, and F-measure. The Precision represents the number of correct concepts found compared to the number of concepts found by the system.

$$Precision = \frac{\#\ correct\ concepts\ found}{\#\ concepts\ found} \qquad (5)$$

The Recall represents the number of correct concepts found by the system with regard to the reference concepts.

$$Recal = \frac{\#correct\ concepts\ found}{\#refernce\ concepts} \qquad (6)$$

The F-measure combines Precision and Recall according to the following equation:

$$F - mesure = \frac{2 \times Recal \times Precision}{Recal + Precision} \qquad (7)$$

We used in our experiments, the free tool CRF++1 for the training and testing steps. It should be noted that the CRF++ tool implements learning by Newtonian method and uses decoding using the Viterbi algorithm.

V. RESULTS AND DISCUSSION

In order to evaluate the performance of CRF models, we used several models that differ according to the level of integration of local dependencies in the same turn. These dependencies vary according to the unigram (one word) or bi-gram (two words) interval of the word to label.

After an initial test phase, we limited the number of models tested to four.

- The first (Model 0) is a model that does not take into account any dependence between the words of the same turn. In this case, CRF models play the role of a simple semantic tagger.

- The second (Model 1) is a model that uses a two-word window taking into account the previous word and the next word in the same turn.

- The third (Model 2) is a model that uses a window involving two words before and two words after the current word.

- The fourth model (Model 3) consists of improving the third model by adding two local dependencies. This dependency uses two bigrams taking into account the current word with the precedent word (respectively with the next word).

Then, we use these models for learning the CRF parameters based on different versions of the annotated corpus (version I,

II and III). The Table VIII, Table IX and Table X illustrate the results of the evaluation of the concept labeling in terms of Precision, Recall, F-measure and Concept Error Rate (CER).

TABLE VIII. RESULTS OF THE CONCEPTUAL LABELING OF VERSION I OF THE CORPUS

		Model 0	Model 1	Model 2	Model 3
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. RESULTS OF THE CONCEPTUAL LABELING OF VERSION II OF THE CORPUS

		Model 0	Model 1	Model 2	Model 3
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. RESULTS OF THE CONCEPTUAL LABELING OF VERSION III OF THE CORPUS

		Model 0	Model 1	Model 2	Model 3
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

Based on these experiments, we notice that the CER decreases with the improvement of the quality of data used for learning. This clearly shows that the pre-processing carried out makes it possible to improve the speech understanding. We noticed also that the models that take into account the dependence between the different words of the same turn (Model 2 and Model 3) make it possible to improve the speech understanding. This is justified by the decrease in the CER and the increase in the F-measure.

Besides these results, we justify the robustness of CRF models with not processed data. Table VIII shows that the F-measure is 77.75% for the "Model 0" which does not take into account underlying dependencies, and reaches 84.40% for the "Model 3" by introducing the bi-model gram.

The examination of the errors made by the CRF models, directed us to carry out other experiments by exploiting the same test corpus, but with considering the dependence of the turn according to the dialogical level and exploiting the sets A, D and X (Table VI). Based on this classification, we tested the CRF models using these different sets on the different versions of preprocessing corpus.

Indeed, we obtained three different versions according to the processing performed for each set A, D and X, starting from the raw version to the fully processed version for each series. The conceptual labeling of the different series is based on the same CRF models based on the different models (Model 0, Model 1, Model 2 and Model 3) presented previously.

Table XI summarizes the results obtained in terms of CER, Precision, Recall and F-measure. From these results, we notice that the CER obtained on the type A speech turns (speech context-independent set) is the lowest rate, comparing it with the results of sets D and X. Therefore, we can conclude that a large part of the errors is due to the presence of out-of-context statements of the dialogue (Set X) and of context-dependent statements (Set D). These results are expected since we are interested in this work in the literal understanding, which does not depend on the dialogical context, so, the turns depending on the context increase the error.

There are other sources contribute to the increase CER. It is mainly about the appearance of terms that are not processed in the training corpus. This is due to the presence of certain phenomena linked to spontaneous speech such as hesitation allowing the addition of out-of-vocabulary words.

TABLE XI. RESULTS OF THE CONCEPTUAL LABELING OF SETS A, D AND X WITH DIFFERENT TYPES OF PROCESSED DATA (VERSION I, II AND III)

			Version I		
			A	D	X
Model 0	CER	(%)	16.93	23.21	38.59
	Precision	(%)	82.58	75.68	59.25
	Recall	(%)	80.32	72.24	56.14
	F-measure	(%)	81.43	73.92	57.65
Model 1	CER	(%)	8.83	12.53	12.28
	Precision	(%)	90.45	85.99	84.09
	Recall	(%)	83.73	76.95	64.91
	F-measure	(%)	86.96	81.22	73.26
Model 2	CER	(%)	8.03	10.26	10.52
	Precision	(%)	91.17	88.15	85.00
	Recall	(%)	82.99	76.36	59.64
	F-measure	(%)	86.89	81.83	70.10
Model 3	CER	(%)	7.29	9.92	10.51
	Precision	(%)	91.87	88.44	86.36
	Recall	(%)	82.46	75.94	66.66
	F-measure	(%)	86.91	81.71	75.24

			Version II		
			A	D	X
Model 0	CER	(%)	15.38	21.76	42.10
	Precision	(%)	84.20	77.36	55.55
	Recall	(%)	82.00	74.37	52.63

	F-measure	(%)	83.09	75.83	54.05

Model 1	CER	(%)	7.02	11.34	17.54
	Precision	(%)	92.55	87.53	78.26
	Recall	(%)	87.35	79.66	63.15
	F-measure	(%)	89.88	83.41	69.90

Model 2	CER	(%)	6.95	10.25	14.03
	Precision	(%)	92.49	88.55	81.39
	Recall	(%)	85.75	79.23	61.40
	F-measure	(%)	88.99	83.68	70.00

Model 3	CER	(%)	6.15	9.66	8.77
	Precision	(%)	93.30	89.13	87.17
	Recall	(%)	85.81	79.24	59.64
	F-measure	(%)	89.40	83.89	70.83

			Version III		
			A	D	X
Model 0	CER	(%)	17.61	22.23	42.10
	Precision	(%)	81.92	76.76	55.55
	Recall	(%)	79.84	73.74	52.63
	F-measure	(%)	80.87	75.22	54.05

Model 1	CER	(%)	6.84	11.95	17.54
	Precision	(%)	92.75	86.97	78.26
	Recall	(%)	87.73	73.74	63.15
	F-measure	(%)	90.17	75.22	69.90

Model 2	CER	(%)	6.65	10.95	10.52
	Precision	(%)	92.83	87.82	85.00
	Recall	(%)	86.23	79.01	59.64
	F-measure	(%)	89.41	83.18	70.10

Model 3	CER	(%)	5.74	10.61	8.77
	Precision	(%)	93.76	88.10	87.17
	Recall	(%)	86.30	78.67	59.64
	F-measure	(%)	89.87	83.12	70.83

In conclusion, CRF models perform well even with not processed turns. On the other hand, conceptual labeling based on CRF models failed when dealing with new terms that are not in the training corpus. These terms can be non-vocabulary words or domain words. This last case is mainly due to the reduced size of the corpus used to learn the CRF models.

VI. CONCLUSION

In this work, we proposed to evaluate in depth the performance of CRF models in the context of speech understanding in dialogue systems. We tested CRF in different models and in different types of processed data. We proved that these models show robustness against noisy data. They recorded good results for conceptual labeling (F-measure of 86.52%). Thus, we found that CRF models have the ability to detect task-specific compound words and label them correctly. These interpretations confirm the performance of these models even for under-resourced languages. As future work, we planify to compare these results with deep learning models for the same task to compare performance between machine learning using CRF models and deep learning models such as CNN (Convolutional Neural Network).

REFERENCES

[1] W. Minker, "Compréhension Automatique de la Parole Spontanée". Éditions de l'Harmattan, 1999.

[2] C. Raymond, and G. Riccardi, "Generative and Discriminative Algorithms for Spoken Language Understanding. In proceeding of Interspeech", pp.1605-1608, 2007.

[3] S. Young, M. Gasic, B. Thomson, and J. D. Williams, "POMDP-Based Statistical Spoken Dialog Systems: A Review". In proceeding of the IEEE, vol. 101, pp.1160 1179, 2013.

[4] C. Lhioui, A. Zouaghi, and M. Zrigui, "A Combined Method Based on Stochastic and Linguistic Paradigm for the Understanding of Arabic Spontaneous Utterances". In proceeding of the international conference on intelligent text processing and computational linguistics (CICLing), lecture notes in computer science (LNCS) springer, vol. 7817, pp.549-558, 2013.

[5] 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.

[6] Y. Y. Wang and A. Acero, "Discriminative Models for Spoken Language Understanding". In proceeding of international conference on spoken language processing (ISCA), pp.1766 1769, 2006.

[7] D. Y. Liliana and C. Basaruddin, "A review on conditional random fields as a sequential classifier in machine learning", 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), Palembang, Indonesia, 2017, pp. 143-148.

[8] Y. Tsuruoka, J. Tsujii, and S. Ananiadou, "Fast Full Parsing By Linear-Chain Conditional Random Fields". In proceeding of the conference of the European chapter of the ACL, pp.790–798, 2009.

[9] I. Tellier, I. Eshkol, S. Taala, and J.P. Prost, "Pos-tagging for oral texts with CRF and category decomposition". Research in computing science, vol. 46, pp.79–90, 2010.

[10] A. Barhoumi, C. Aloulou, I. Zitouni, and L.H. Belguith," Analyse Syntaxique Statistique de la Langue Arabe ". (CEC-TAL 2015) pp.38-46, 2015.

[11] Y.Y. Wang, A. Acero, M. Mahajan, J. Lee, "Combining Statistical and Knowledge-Based Spoken Language Understanding in Conditional Models". In proceeding of the international conference on computational linguistics and the annual meeting of the association for computational linguistics COLING-ACL, pp.882-889, 2006.

[12] S. Hahn, M. Dinarelli, C. Raymond, F. Lefèvre, O. Lehnen, R. De Mori, A. Moschitti, H. Ney, and G. Riccardi, "Comparing Stochastic Approaches to Spoken Language Understanding in Multiple Languages". IEEE transaction on audio, speech, and language processing, vol. 19, pp.1569-1583, 2011.

[13] A. Deoras, G. Tur, R. Sarikaya, and D. Hakkani-Tür, "Joint Discriminative Decoding of Words and Semantic Tags for Spoken

Language Understanding". IEEE transactions on audio, speech, and language processing, vol. 21, pp.1612-1621, 2013.

[14] J. Lafferty, A. McCallum, F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data". In proceeding of the international conference on machine learning (ICML), pp.282-289, 2001.

[15] M. Graja, M. Jaoua, and L.H. Belguith, "Statistical Framework with Knowledge Base Integration for Robust Speech Understanding of the Tunisian Dialect". IEEE Transactions on Audio, Speech, and Language Processing (T-ASLP), 2015.

[16] M. Graja, M. Jaoua, and L.H. Belguith, "Discriminative Framework for Spoken Tunisian Dialect Understanding". In proceeding of international conference on statistical language and speech processing (SLSP).

Lecture notes in computer science, LNCS, vol. 7978, pp.102-110, Springer, 2013.

[17] W. Minker and S. Bennacef S, "Speech and Human Machine Dialog". The springer international series in engineering and computer science, kluwer academic publishers, vol. 770, 2004.

[18] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields. Foundations and trends in machine learning", vol. 4, pp.267–373, 2012.

[19] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning", MIT Press, 2006.

[20] T. Do and T. Artières, " Champs de Markov Conditionnels pour le Traitement de Séquences ". Revue des nouvelles technologies de l'information, extraction et gestion de connaissances, vol. 2, pp.639-650, 2006.