Semantic Role Labeling Systems for Arabic using Kernel Methods

Mona Diab
CCLS, Columbia University
New York, NY 10115, USA
mdiab@ccls.columbia.edu

Alessandro Moschitti
DISI, University of Trento
Trento, I-38100, Italy
moschitti@disi.unitn.it

Daniele Pighin
FBK-irst; DISI, University of Trento
Trento, I-38100, Italy
pighin@fbk.eu

Abstract

There is a widely held belief in the natural language and computational linguistics communities that Semantic Role Labeling (SRL) is a significant step toward improving important applications, e.g. question answering and information extraction. In this paper, we present an SRL system for Modern Standard Arabic that exploits many aspects of the rich morphological features of the language. The experiments on the pilot Arabic Propbank data show that our system based on Support Vector Machines and Kernel Methods yields a global SRL F_1 score of 82.17%, which improves the current state-of-the-art in Arabic SRL.

1 Introduction

Shallow approaches to semantic processing are making large strides in the direction of efficiently and effectively deriving tacit semantic information from text. Semantic Role Labeling (SRL) is one such approach. With the advent of faster and more powerful computers, more effective machine learning algorithms, and importantly, large data resources annotated with relevant levels of semantic information, such as the FrameNet (Baker et al., 1998) and ProbBank (Kingsbury and Palmer, 2003), we are seeing a surge in efficient approaches to SRL (Carreras and Màrquez, 2005).

SRL is the process by which predicates and their arguments are identified and their roles are defined in a sentence. For example, in the English sentence, 'John likes apples.', the predicate is 'likes' whereas 'John' and 'apples', bear the semantic role labels *agent* (ARG0) and *theme* (ARG1). The crucial fact about semantic roles is that regardless of the overt syntactic structure variation, the underlying predicates remain the same. Hence, for the sentence 'John opened the door' and 'the door opened', though 'the door' is the object of the first sentence

and the subject of the second, it is the 'theme' in both sentences. Same idea applies to passive constructions, for example.

There is a widely held belief in the NLP and computational linguistics communities that identifying and defining roles of predicate arguments in a sentence has a lot of potential for and is a significant step toward improving important applications such as document retrieval, machine translation, question answering and information extraction (Moschitti et al., 2007).

To date, most of the reported SRL systems are for English, and most of the data resources exist for English. We do see some headway for other languages such as German and Chinese (Erk and Pado, 2006; Sun and Jurafsky, 2004). The systems for the other languages follow the successful models devised for English, e.g. (Gildea and Jurafsky, 2002; Gildea and Palmer, 2002; Chen and Rambow, 2003; Thompson et al., 2003; Pradhan et al., 2003; Moschitti, 2004; Xue and Palmer, 2004; Haghighi et al., 2005). In the same spirit and facilitated by the release of the SemEval 2007 Task 18 data[1], based on the Pilot Arabic Propbank, a preliminary SRL system exists for Arabic[2] (Diab and Moschitti, 2007; Diab et al., 2007a). However, it did not exploit some special characteristics of the Arabic language on the SRL task.

In this paper, we present an SRL system for MSA that exploits many aspects of the rich morphological features of the language. It is based on a supervised model that uses support vector machines (SVM) technology (Vapnik, 1998) for argument boundary detection and argument classification. It is trained and tested using the pilot Arabic Propbank data released as part of the SemEval 2007 data. Given the lack of a reliable Arabic deep syntactic parser, we

---

[1]http://nlp.cs.swarthmore.edu/semeval/
[2]We use Arabic to refer to Modern Standard Arabic (MSA).

use gold standard trees from the Arabic Tree Bank (ATB) (Maamouri et al., 2004).

This paper is laid out as follows: Section 2 presents facts about the Arabic language especially in relevant contrast to English; Section 3 presents the approach and system adopted for this work; Section 4 presents the experimental setup, results and discussion. Finally, Section 5 draws our conclusions.

2   Arabic Language and Impact on SRL

Arabic is a very different language from English in several respects relevant to the SRL task. Arabic is a semitic language. It is known for its templatic morphology where words are made up of roots and affixes. Clitics agglutinate to words. Clitics include prepositions, conjunctions, and pronouns.

In contrast to English, Arabic exhibits rich morphology. Similar to English, Arabic verbs explicitly encode tense, voice, Number, and Person features. Additionally, Arabic encodes verbs with Gender, Mood (subjunctive, indicative and jussive) information. For nominals (nouns, adjectives, proper names), Arabic encodes syntactic Case (accusative, genitive and nominative), Number, Gender and Definiteness features. In general, many of the morphological features of the language are expressed via short vowels also known as diacritics[3].

Unlike English, syntactically Arabic is a pro-drop language, where the subject of a verb may be implicitly encoded in the verb morphology. Hence, we observe sentences such as اكل البرتقال *Akl AlbrtqAl* 'ate-[he] the-oranges', where the verb *Akl* encodes the third Person Masculine Singular subject in the verbal morphology. It is worth noting that in the ATB 35% of all sentences are pro-dropped for subject (Maamouri et al., 2006). Unless the syntactic parse is very accurate in identifying the pro-dropped case, identifying the syntactic subject and the underlying semantic arguments are a challenge for such pro-drop cases.

Arabic syntax exhibits relative free word order. Arabic allows for both subject-verb-object (SVO) and verb-subject-object (VSO) argument orders.[4] In the VSO constructions, the verb agrees with the syntactic subject in Gender only, while in the SVO constructions, the verb agrees with the subject in both Number and Gender. Even though, in the ATB, an equal distribution of both VSO and SVO is observed (each appearing 30% of the time), it is known that in general Arabic is predominantly in VSO order. Moreover, the pro-drop cases could effectively be perceived as VSO orders for the purposes of SRL. Syntactic Case is very important in the cases of VSO and pro-drop constructions as they indicate the syntactic roles of the object arguments with accusative Case. Unless the morphology of syntactic Case is explicitly present, such free word order could run the SRL system into significant confusion for many of the predicates where both arguments are semantically of the same type.

Arabic exhibits more complex noun phrases than English mainly to express possession. These constructions are known as *idafa* constructions. Modern standard Arabic does not have a special particle expressing possession. In these complex structures a surface indefinite noun (missing an explicit definite article) may be followed by a definite noun marked with genitive Case, rendering the first noun syntactically definite. For example, رجل البيت *rjl Albyt* 'man the-house' meaning 'man of the house', رجل becomes definite. An adjective modifying the noun رجل will have to agree with it in Number, Gender, Definiteness, and Case. However, without explicit morphological encoding of these agreements, the scope of the arguments would be confusing to an SRL system. In a sentence such as رجل البيت الطويل *rjlu Albyti AlTwylu* meaning 'the tall man of the house': 'man' is definite, masculine, singular, nominative, corresponding to Definiteness, Gender, Number and Case, respectively; 'the-house' is definite, masculine, singular, genitive; 'the-tall' is definite, masculine, singular, nominative. We note that 'man' and 'tall' agree in Number, Gender, Case and Definiteness. Syntactic Case is marked using short vowels *u*, and *i* at the end of the word. Hence, *rjlu* and *AlTwylu* agree in their Case ending[5] Without the explicit marking of the Case information,

---

[3]Diacritics encode the vocalic structure, namely the short vowels, as well as the gemmination marker for consonantal doubling, among other markers.

[4]MSA less often allows for OSV, or OVS.

[5]The presence of the *Albyti* is crucial as it renders *rjlu* definite therefore allowing the agreement with *AlTwylu* to be complete.
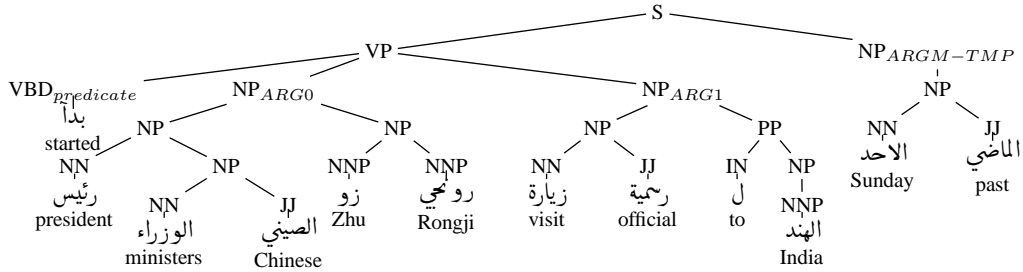
S

VP

VBD$_{predicate}$ NP$_{ARG0}$ NP$_{ARG1}$ NP$_{ARGM-TMP}$

بدأ
started

NP   NP   NP   PP   NP

NN
رئيس
president

NN   JJ   NNP   NNP   NN   JJ   IN   NP   NN   JJ

الوزراء
ministers

الصيني
Chinese

زو
Zhu

روﻨﺠﻲ
Rongji

زيارة
visit

رسمية
official

ل
to

NNP
الهند
India

الاحد
Sunday

الماضي
past

Figure 1: Annotated Arabic Tree corresponding to 'Chinese Prime minister Zhu Rongjy started an official visit to India last Sunday.'

namely in the word endings, it could be equally valid that 'the-tall' modifies 'the-house' since they agree in Number, Gender and Definiteness as explicitly marked by the Definiteness article *Al*. Hence, these idafa constructions could be tricky for SRL in the absence of explicit morphological features. This is compounded by the general absence of short vowels, expressed by diacritics (i.e. the *u* and *i* in *rjlu* and *Al-byti*,) in naturally occurring text. Idafa constructions in the ATB exhibit recursive structure, embedding other NPs, compared to English where possession is annotated with flat NPs and is designated by a possessive marker.

Arabic texts are underspecified for diacritics to different degrees depending on the genre of the text (Diab et al., 2007b). Such an underspecification of diacritics masks some of the very relevant morpho-syntactic interactions between the different categories such as agreement between nominals and their modifiers as exemplified before, or verbs and their subjects.

Having highlighted the differences, we hypothesize that the interaction between the rich morphology (if explicitly marked and present) and syntax could help with the SRL task. The presence of explicit Number and Gender agreement as well as Case information aids with identification of the syntactic subject and object even if the word order is relatively free. Gender, Number, Definiteness and Case agreement between nouns and their modifiers and other nominals, should give clues to the scope of arguments as well as their classes. The presence of such morpho-syntactic information should lead to better argument boundary detection and better classification.

3 An SRL system for Arabic

The previous section suggests that an optimal model should take into account specific characteristics of

Feature Name	Description
Predicate	Lemmatization of the predicate word
Path	Syntactic path linking the predicate and an argument, e.g. NN↑NP↑VP↓VBX
Partial path	*Path* feature limited to the branching of the argument
No-direction path	Like *Path* without traversal directions
Phrase type	Syntactic type of the argument node
Position	Relative position of the argument with respect to the predicate
Verb subcategorization	Production rule expanding the predicate parent node
Syntactic Frame	Position of the NPs surrounding the predicate
First and last word/POS	First and last words and POS tags of candidate argument phrases

Table 1: Standard linguistic features employed by most SRL systems.

Arabic. In this research, we go beyond the previously proposed basic SRL system for Arabic (Diab et al., 2007a; Diab and Moschitti, 2007). We exploit the full morphological potential of the language to verify our hypothesis that taking advantage of the interaction between morphology and syntax can improve on a basic SRL system for morphologically rich languages.

Similar to the previous Arabic SRL systems, our adopted SRL models use Support Vector Machines to implement a two step classification approach, i.e. boundary detection and argument classification. Such models have already been investigated in (Pradhan et al., 2005; Moschitti et al., 2005). The two step classification description is as follows.

3.1 Predicate Argument Extraction

The extraction of predicative structures is based on the sentence level. Given a sentence, its predicates, as indicated by verbs, have to be identified along with their arguments. This problem is usually divided in two subtasks: (a) the detection of the target argument boundaries, i.e. the span of the argument words in the sentence, and (b) the classification of the argument type, e.g. *Arg0* or *ArgM* for Propbank
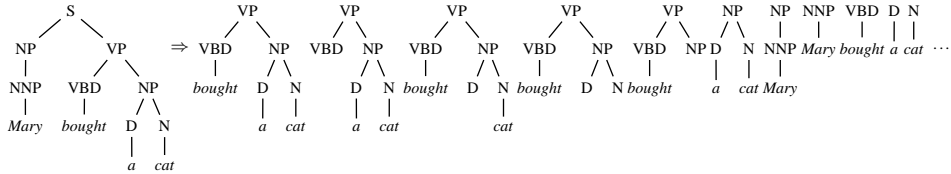
Figure 2: Fragment space generated by a tree kernel function for the sentence *Mary bought a cat*.

or *Agent* and *Goal* for the FrameNet.

The standard approach to learn both the detection and the classification of predicate arguments is summarized by the following steps:
(a) Given a sentence from the *training-set*, generate a full syntactic parse-tree;
(b) let $\mathcal{P}$ and $\mathcal{A}$ be the set of predicates and the set of parse-tree nodes (i.e. the potential arguments), respectively;
(c) for each pair $\langle p, a \rangle \in \mathcal{P} \times \mathcal{A}$: extract the feature representation set, $F_{p,a}$ and put it in T^+ (positive examples) if the subtree rooted in a covers exactly the words of one argument of p, otherwise put it in T^- (negative examples).

For instance, in Figure 1, for each combination of the predicate *started* with the nodes NP, S, VP, VPD, NNP, NN, PP, JJ or IN the instances $F_{started,a}$ are generated. In case the node a exactly covers 'president ministers Chinese Zhu Rongji' or 'visit official to India', $F_{p,a}$ will be a positive instance otherwise it will be a negative one, e.g. $F_{started,IN}$.

The T^+ and T^- sets are used to train the boundary classifier. To train the multi-class classifier, T^+ can be reorganized as positive $T_{arg_i}^+$ and negative $T_{arg_i}^-$ examples for each argument i. This way, an individual ONE-vs-ALL classifier for each argument i can be trained. We adopt this solution, according to (Pradhan et al., 2005), since it is simple and effective. In the classification phase, given an unseen sentence, all its $F_{p,a}$ are generated and classified by each individual classifier C_i. The argument associated with the maximum among the scores provided by the individual classifiers is eventually selected.

The above approach assigns labels independently, without considering the whole predicate argument structure. As a consequence, the classifier output may generate overlapping arguments. Thus, to make the annotations globally consistent, we apply a disambiguating heuristic adopted from (Diab and Moschitti, 2007) that selects only one argument among multiple overlapping arguments.

3.2 Features

The discovery of relevant features is, as usual, a complex task. The choice of features is further compounded for a language such as Arabic given its rich morphology and morpho-syntactic interactions.

To date, there is a common consensus on the set of basic standard features for SRL, which we will refer to as *standard*. The set of standard features, refers to unstructured information derived from parse trees. e.g. *Phrase Type*, *Predicate Word* or *Head Word*. Typically the standard features are language independent. In our experiments we employ the features listed in Table 1, defined in (Gildea and Jurafsky, 2002; Pradhan et al., 2005; Xue and Palmer, 2004). For example, the *Phrase Type* indicates the syntactic type of the phrase labeled as a predicate argument, e.g. NP for *ARG1* in Figure 1. The *Parse Tree Path* contains the path in the parse tree between the predicate and the argument phrase, expressed as a sequence of nonterminal labels linked by direction (up or down) symbols, e.g. VBD ↑ VP ↓ NP for *ARG1* in Figure 1. The *Predicate Word* is the surface form of the verbal predicate, e.g. *started* for all arguments. The standard features, as successful as they are, are designed primarily for English. They are not exploiting the different characteristics of the Arabic language as expressed through morphology. Hence, we explicitly encode new SRL features that capture the richness of Arabic morphology and its role in morpho-syntactic behavior. The set of morphological attributes include: inflectional morphology such as Number, Gender, Definiteness, Mood, Case, Person; derivational morphology such as the Lemma form of the words with all the diacritics explicitly marked; vowelized and fully diacritized form of the surface form; the English gloss[6]. It is worth noting that there exists highly accurate morphological taggers for Arabic such as the MADA system (Habash and Rambow, 2005; Roth et al., 2008). MADA tags

---

[6]The gloss is not sense disambiguated, hence they include homonyms.

Feature Name	Description
Definiteness	Applies to nominals, values are definite, indefinite or inapplicable
Number	Applies to nominals and verbs, values are singular, plural or dual or inapplicable
Gender	Applies to nominals, values are feminine, masculine or inapplicable
Case	Applies to nominals, values are accusative, genitive, nominative or inapplicable
Mood	Applies to verbs, values are subjunctive, indicative, jussive or inapplicable
Person	Applies to verbs and pronouns, values are 1st, 2nd, 3rd person or inapplicable
Lemma	The citation form of the word fully diacritized with the short vowels and gemmination markers if applicable
Gloss	this is the corresponding English meaning as rendered by the underlying lexicon.
Vocalized word	The surface form of the word with all the relevant diacritics. Unlike Lemma, it includes all the inflections.
Unvowelized word	The naturally occurring form of the word in the sentence with no diacritics.

Table 2: Rich morphological features encoded in the Extended Argument Structure Tree (EAST).

modern standard Arabic with all the relevant morphological features as well as it produces highly accurate lemma and gloss information by tapping into an underlying morphological lexicon. A list of the extended features is described in Table 2.

The set of possible features and their combinations are very large leading to an intractable feature selection problem. Therefore, we exploit well known kernel methods, namely tree kernels, to robustly experiment with all the features simultaneously. Such kernel engineering, as shown in (Moschitti, 2004), allows us to experiment with many syntactic/semantic features seamlessly.

3.3 Engineering Arabic Features with Kernel Methods

Feature engineering via kernel methods is a useful technique that allows us to save a lot of time in the design and implementation of features. The basic idea is (a) to design a set of basic value-attribute features and apply polynomial kernels and generate all possible combinations; or (b) to design basic tree structures expressing properties related to the target linguistic objects and use tree kernels to generate all possible tree subparts, which will constitute the feature representation vectors for the learning algorithm.

Tree kernels evaluate the similarity between two trees in terms of their overlap, generally measured as the number of common substructures (Collins and Duffy, 2002). For example, Figure 2, shows a small parse tree and some of its fragments. To design a function which computes the number of common substructures between two trees t_1 and t_2, let us define the set of fragments $\mathcal{F}=\{f_1, f_2, ..\}$ and the indicator function $I_i(n)$, equal to 1 if the target f_i is rooted at node n and 0 otherwise. A tree kernel function $K_T(\cdot)$ over two trees is defined as:
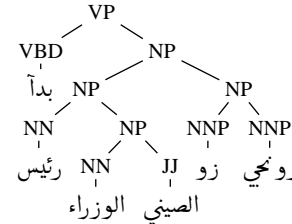

Figure 3: Example of the positive AST structured feature encoding the argument ARG0 in the sentence depicted in Figure 1.

$K_T(t_1, t_2) = \sum_{n_1 \in N_{t_1}} \sum_{n_2 \in N_{t_2}} \Delta(n_1, n_2)$, where N_{t_1} and N_{t_2} are the sets of nodes of t_1 and t_2, respectively. The function $\Delta(\cdot)$ evaluates the number of common fragments rooted in n_1 and n_2, i.e. $\Delta(n_1, n_2) = \sum_{i=1}^{|\mathcal{F}|} I_i(n_1)I_i(n_2)$. Δ can be efficiently computed with the algorithm proposed in (Collins and Duffy, 2002).

3.4 Structural Features for Arabic

In order to incorporate the characteristically rich Arabic morphology features structurally in the tree representations, we convert the features into value-attribute pairs at the leaf node level of the tree. Fig 1 illustrates the morphologically underspecified tree with some of the morphological features encoded in the POS tag such as VBD indicating past tense. This contrasts with Fig. 4 which shows an excerpt of the same tree encoding the chosen relevant morphological features.

For the sake of classification, we will be dealing with two kinds of structures: the Argument Structure Tree (AST) (Pighin and Basili, 2006) and the Extended Argument Structure Tree (EAST). The AST is defined as the minimal subtree encompassing all and only the leaf nodes encoding words belonging to the predicate or one of its arguments. An AST example is shown in Figure 3. The EAST is the corresponding structure in which all the leaf nodes have been extended with the ten morphological fea-
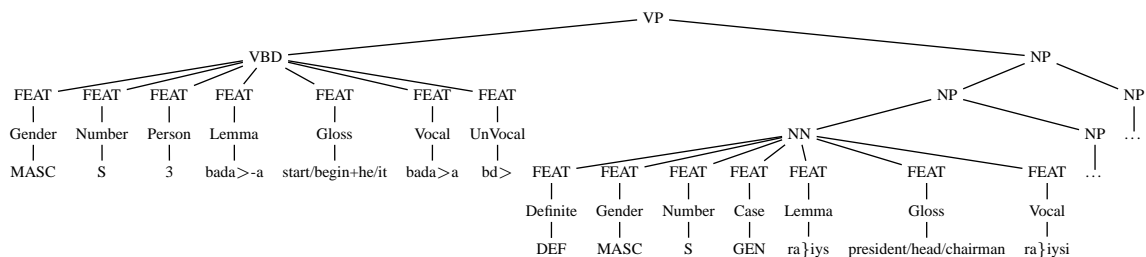
VP

VBD ... NP

FEAT FEAT FEAT FEAT FEAT FEAT FEAT ... NP ... NP

Gender Number Person Lemma Gloss Vocal UnVocal ... NN ... NP ...

MASC S 3 bada>-a start/begin+he/it bada>a bd>

FEAT FEAT FEAT FEAT FEAT FEAT FEAT ...

Definite Gender Number Case Lemma Gloss Vocal

DEF MASC S GEN ra}iys president/head/chairman ra}iysi

Figure 4: An excerpt of the EAST corresponding to the AST shown in Figure 3, with attribute-value extended morphological features represented as leaf nodes.

tures described in Table 2, forming a vector of 10 preterminal-terminal node pairs that replace the surface of the leaf. The resulting EAST structure is shown in Figure 4.

Not all the features are instantiated for all the leaf node words. Due to space limitations, in the figure we did not include the Features that have NULL values. For instance, Definiteness is always associated with nominals, hence the verb بدآ *bd'* 'start' is assigned a NULL value for the Definite feature. Verbs exhibit Gender information depending on inflections. For our example, بدآ 'started' is inflected for masculine Gender, singular Number, third person. On the other hand, the noun الوزراء is definite and is assigned genitive Case since it is in a possessive, idafa, construction.

The features encoded by the EAST can provide very useful hints for boundary and role classification. Considering Figure 1, argument boundaries is not as straight forward to identify as there are several NPs. Assuming that the inner most NP 'ministers the-Chinese' is a valid Argument could potentially be accepted. There is ample evidence that any NN followed by a JJ would make a perfectly valid Argument. However, an AST structure would mask the fact that the JJ 'the-Chinese' does not modify the NN 'ministers' since they do not agree in Number[7], and in syntactic Case, where the latter is genitive and the former is nominative. 'the-Chinese' in fact modifies 'president' as they agree on all the underlying morphological features. Conversely, the EAST in Figure 4 explicitly encodes this agreement including an agreement on Definiteness. It is worth noting that just observing the Arabic word رئيس 'president' in Fig 1, the system would assume that it is an indefinite word since it does not include the definite arti-

---

[7]The POS tag on this node is NN as broken plural, however, the underlying morphological feature Number is plural.

cle ال. Therefore, the system could be lead astray to conclude that 'the-Chinese' does not modify 'president' but rather 'the-ministers'. Without knowing the Case information and the agreement features between the verb بدآ 'started' and the two nouns heading the two main NPs in our tree, the syntactic subject can be either زيارة 'visit' or رئيس 'president' in Figure 1. The EAST is more effective in identifying the first noun as the syntactic subject and the second as the object since the morphological information indicates that they are in nominative and accusative Case, respectively. Also the agreement in Gender and Number between the verb and the syntactic subject is identified in the enriched tree. We see that بدآ 'started' and رئيس 'president' agree in being singular and masculine. If زيارة 'visit' were the syntactic subject, we would have seen the verb inflected as بدأت 'started-FEM' with a feminine inflection to reflect the verb-subject agreement on Gender. Hence these agreement features should help with the classification task.

4 Experiments

In these experiments we investigate (a) if the technology proposed in previous work for automatic SRL of English texts is suitable for Arabic SRL systems, and (b) the impact of tree kernels using new tree structures on Arabic SRL. For this purpose, we test our models on the two individual phases of the traditional 2-stage SRL model (i.e. boundary detection and argument classification) and on the complete SRL task. We use three different feature spaces: a set of standard attribute-value features and the AST and the EAST structures defined in 3.4. Standard feature vectors can be combined with a polynomial kernel (Poly), which, when the degree is larger than 1, automatically generates feature conjunctions. This, as suggested in (Pradhan et al., 2005; Moschitti, 2004), can help stressing the differ-

ences between different argument types. Tree structures can be used in the learning algorithm thanks to the tree kernels described in Section 3.3. Moreover, to verify if the above feature sets are equivalent or complementary, we can join them by means of additive operation which always produces a valid kernel (Shawe-Taylor and Cristianini, 2004).

4.1 Experimental setup

We use the dataset released in the SemEval 2007 Task 18 on Arabic Semantic Labeling (Diab et al., 2007a). The data covers the 95 most frequent verbs in the Arabic Treebank III ver. 2 (ATB). The ATB consists of MSA newswire data from the Annhar newspaper, spanning the months from July to November, 2002. All our experiments are carried out with gold standard trees.

An important characteristic of the dataset is the use of unvowelized Arabic in the Buckwalter transliteration scheme for deriving the basic features for the AST experimental condition. The data comprises a development set, a test set and a training set of 886, 902 and 8,402 sentences, respectively, where each set contain 1725, 1661 and 21,194 argument instances. These instances are distributed over 26 different role types. The training instances of the boundary detection task also include parse-tree nodes that do not correspond to correct boundaries (we only considered 350K examples). For the experiments, we use SVM-Light-TK toolkit[8] (Moschitti, 2004; Moschitti, 2006) and its SVM-Light default parameters. The system performance, i.e. F_1 on single boundary and role classifier, accuracy of the role multi-classifier and the F_1 of the complete SRL systems, are computed by means of the CoNLL evaluator[9].

4.2 Results

Figure 5 reports the F_1 of the SVM boundary classifier using Polynomial Kernels with a degree from 1 to 6 (i.e. Poly$i$), the AST and the EAST kernels and their combinations. We note that as we introduce conjunctions, i.e. a degree larger than 2, the F_1 increases by more than 3 percentage points. Thus, not only are the English features meaningful for Arabic but also their combinations are important, reveal-

[8]http://disi.unitn.it/∼moschitti
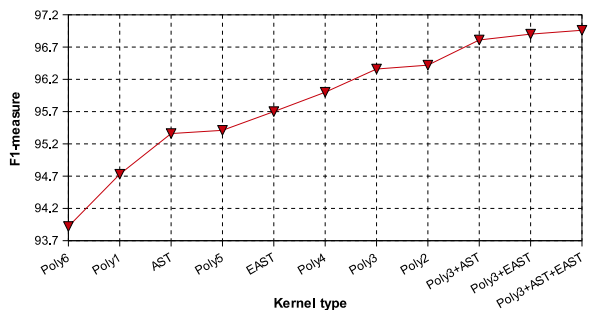[9]http://www.lsi.upc.es/∼srlconll/soft.html

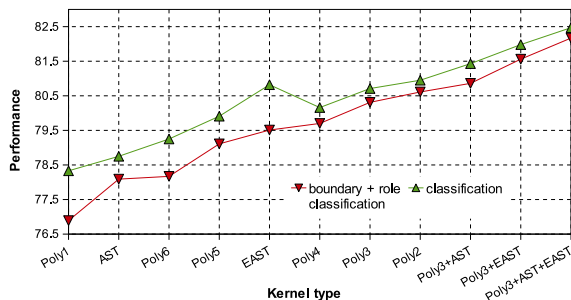Figure 5: Impact of polynomial kernel, tree kernels and their combinations on boundary detection.

Figure 6: Impact of the polynomial kernel, tree kernels and their combinations on the accuracy in role classification (gold boundaries) and on the F1 of complete SRL task (boundary + role classification).

ing that both languages share an underlying syntax-semantics interface. Moreover, we note that the F_1 of EAST is higher than the F_1 of AST which in turn is higher than the linear kernel (Poly1). However, when conjunctive features (Poly2-4) are used the system accuracy exceeds those of tree kernel models alone. Further increasing the polynomial degree (Poly5-6) generates very complex hypotheses which result in very low accuracy values.

Therefore, to improve the polynomial kernel, we sum it to the contribution of AST and/or EAST, obtaining AST+Poly3 (polynomial kernel of degree 3), EAST+Poly3 and AST+EAST+Poly3, whose F_1 scores are also shown in Figure 5. Such combined models improve on the best polynomial kernel. However, not much difference is shown between AST and EAST on boundary detection. This is expected since we are using gold standard trees. We hypothesize that the rich morphological features will help more with the role classification task. Therefore, we evaluate role classification with gold boundaries. The curve labeled "classification" in Figure 6 illustrates the accuracy of the SVM role multi-classifier according to different kernels.

804

	P3	AST	EAST	AST+ P3	EAST+ P3	AST+ EAST+ P3
P	81.73	80.33	81.7	81.73	82.46	83.08
R	78.93	75.98	77.42	80.01	80.67	81.28
F_1	80.31	78.09	79.51	80.86	81.56	82.17

Table 3: F_1 of different models on the Arabic SRL task.

Again, we note that a degree larger than 1 yields a significant improvement of more than 3 percent points, suggesting that the design of Arabic SRL system based on SVMs requires polynomial kernels. In contrast to the boundary results, EAST highly improves over AST (by about 3 percentage points) and produces an F_1 comparable to the best Polynomial kernel. Moreover, AST+Poly3, EAST+Poly3 and AST+EAST+Poly3 all yield different degrees of improvement, where the latter model is both the richest in terms of features and the most accurate.

These results strongly suggest that: (a) tree kernels generate new syntactic features that are useful for the classification of Arabic semantic roles; (b) the richer morphology of Arabic language should be exploited effectively to obtain accurate SRL systems; (c) tree kernels appears to be a viable approach to effectively achieve this goal.

To illustrate the practical feasibility of our system, we investigate the complete SRL task where both the boundary detection and argument role classification are performed automatically. The curve labeled "boundary + role classification" in Figure 6 reports the F_1 of SRL systems based on the previous kernels. The trend of the plot is similar to the gold-standard boundaries case. The difference among the F_1 scores of the AST+Poly3, EAST+Poly3 and AST+EAST+Poly3 is slightly reduced. This may be attributed to the fact that they produce similar boundary detection results, which in turn, for the global SRL outcome, are summed to those of the classification phase. Table 3 details the differences among the models and shows that the best model improves the SRL system based on the polynomial kernel, i.e. the SRL state-of-the-art for Arabic, by about 2 percentage points. This is a very large improvement for SRL systems (Carreras and Màrquez, 2005). These results confirm that the new enriched structures along with tree kernels are a promising approach for Arabic SRL systems.

Finally, Table 4 reports the F_1 of the best model, AST+EAST+Poly3, for individual arguments in the

Role	Precision	Recall	$F_{\beta=1}$
ARG0	96.14%	97.27%	96.70
ARG0-STR	100.00%	20.00%	33.33
ARG1	88.52%	92.70%	90.57
ARG1-STR	33.33%	15.38%	21.05
ARG2	69.35%	76.67%	72.82
ARG3	66.67%	16.67%	26.67
ARGM-ADV	66.98%	61.74%	64.25
ARGM-CAU	100.00%	9.09%	16.67
ARGM-CND	25.00%	33.33%	28.57
ARGM-LOC	67.44%	95.08%	78.91
ARGM-MNR	54.00%	49.09%	51.43
ARGM-NEG	80.85%	97.44%	88.37
ARGM-PRD	20.00%	8.33%	11.76
ARGM-PRP	85.71%	66.67%	75.00
ARGM-TMP	91.35%	88.79%	90.05

Table 4: SRL F_1 of the single arguments using the AST+EAST+Poly3 kernel.

SRL task. We note that, as for English SRL, ARG0 shows high values (96.70%). Conversely, ARG1 seems more difficult to be classified in Arabic. The F_1 for ARG1 is only 90.57% compared with 96.70% for ARG0.

This may be attributed to the different possible syntactic orders of Arabic consructions confusing the syntactic subject with the object especially where there is no clear morphological features on the arguments to decide either way.

5 Conclusions

We have presented a model for Arabic SRL that yields a global SRL F_1 score of 82.17% by combining rich structured features and traditional attribute-value features derived from English SRL systems. The resulting system significantly improves previously reported results on the same task and dataset. This outcome is very promising given that the available data is small compared to the English data sets.

For future work, we would like to explore further explicit morphological features such as aspect tense and voice as well as richer POS tag sets such as those proposed in (Diab, 2007). Finally, we would like to experiment with automatic parses and different syntactic formalisms such as dependencies and shallow parses.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *COLING-ACL '98: University of Montréal*.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. In *Proceedings of CoNLL-2005*, Ann Arbor, Michigan.

John Chen and Owen Rambow. 2003. Use of Deep Linguistic Features for the Recognition and Labeling of Semantic Arguments. In *Proceedings of EMNLP*, Sapporo, Japan.

Michael Collins and Nigel Duffy. 2002. New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete structures, and the voted perceptron. In *ACL02*.

Mona Diab and Alessandro Moschitti. 2007. Semantic Parsing for Modern Standard Arabic. In *Proceedings of RANLP*, Borovets, Bulgaria.

Mona Diab, Musa Alkhalifa, Sabry ElKateb, Christiane Fellbaum, Aous Mansouri, and Martha Palmer. 2007a. Semeval-2007 task 18: Arabic Semantic Labeling. In *Proceedings of SemEval-2007*, Prague, Czech Republic.

Mona Diab, Mahmoud Ghoneim, and Nizar Habash. 2007b. Arabic Diacritization in the Context of Statistical Machine Translation. In *Proceedings of MT-Summit*, Copenhagen, Denmark.

Mona Diab. 2007. Towards an Optimal Pos Tag Set for Modern Standard Arabic Processing. In *Proceedings of RANLP*, Borovets, Bulgaria.

Katrin Erk and Sebastian Pado. 2006. Shalmaneser – A Toolchain for Shallow Semantic Parsing. Proceedings of LREC.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*.

Daniel Gildea and Martha Palmer. 2002. The Necessity of Parsing for Predicate Argument Recognition. In *Proceedings of ACL-02*, Philadelphia, PA, USA.

Nizar Habash and Owen Rambow. 2005. Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop. In *Proceedings of ACL'05*, Ann Arbor, Michigan.

Aria Haghighi, Kristina Toutanova, and Christopher Manning. 2005. A Joint Model for Semantic Role Labeling. In *Proceedings ofCoNLL-2005*, Ann Arbor, Michigan.

Paul Kingsbury and Martha Palmer. 2003. Propbank: the Next Level of Treebank. In *Proceedings of Treebanks and Lexical Theories*.

Mohamed Maamouri, Ann Bies, Tim Buckwalter, and Wigdan Mekki. 2004. The Penn Arabic Treebank : Building a Large-Scale Annotated Arabic Corpus.

Mohamed Maamouri, Ann Bies, Tim Buckwalter, Mona Diab, Nizar Habash, Owen Rambow, and Dalila Tabessi. 2006. Developing and Using a Pilot Dialectal Arabic Treebank.

Alessandro Moschitti, Ana-Maria Giuglea, Bonaventura Coppola, and Roberto Basili. 2005. Hierarchical Semantic Role Labeling. In *Proceedings of CoNLL-2005*, Ann Arbor, Michigan.

Alessandro Moschitti, Silvia Quarteroni, Roberto Basili, and Suresh Manandhar. 2007. Exploiting Syntactic and Shallow Semantic Kernels for Question Answer Classification. In *Proceedings of ACL'07*, Prague, Czech Republic.

Alessandro Moschitti. 2004. A Study on Convolution Kernels for Shallow Semantic Parsing. In *proceedings of ACL'04*, Barcelona, Spain.

Alessandro Moschitti. 2006. Making Tree Kernels Practical for Natural Language Learning. In *Proceedings of EACL'06*.

Alessandro Moschitti, Daniele Pighin and Roberto Basili. 2006. Semantic Role Labeling via Tree Kernel Joint Inference. In *Proceedings of CoNLL-X*.

Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2003. Semantic Role Parsing: Adding Semantic Structure to Unstructured Text. In *Proceedings ICDM'03*, Melbourne, USA.

Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2005. Support Vector Learning for Semantic Argument Classification. *Machine Learning*.

Ryan Roth, Owen Rambow, Nizar Habash, Mona Diab, and Cynthia Rudin. 2008. Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking. In *ACL'08, Short Papers*, Columbus, Ohio, June.

John Shawe-Taylor and Nello Cristianini. 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Honglin Sun and Daniel Jurafsky. 2004. Shallow Semantic Parsing of Chinese. In Proceedings of NAACL-HLT.

Cynthia A. Thompson, Roger Levy, and Christopher Manning. 2003. A Generative Model for Semantic Role Labeling. In *ECML'03*.

Vladimir N. Vapnik. 1998. *Statistical Learning Theory*. John Wiley and Sons.

Nianwen Xue and Martha Palmer. 2004. Calibrating Features for Semantic Role Labeling. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, Barcelona, Spain.