International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:6, 2008

Thematic Role Extraction Using Shallow Parsing

Mehrnoush Shamsfard, Maryam Sadr Mousavi

***Abstract*—**Extracting thematic (semantic) roles is one of the major steps in representing text meaning. It refers to finding the semantic relations between a predicate and syntactic constituents in a sentence. In this paper we present a rule-based approach to extract semantic roles from Persian sentences. The system exploits a two-phase architecture to (1) identify the arguments and (2) label them for each predicate.

For the first phase we developed a rule based shallow parser to chunk Persian sentences and for the second phase we developed a knowledge-based system to assign 16 selected thematic roles to the chunks. The experimental results of testing each phase are shown at the end of the paper.

***Keywords*—**Natural Language Processing, Semantic Role Labeling, Shallow parsing, Thematic Roles.

I. INTRODUCTION

IN recent years there has been an increasing interest in semantic parsing of natural language, which is becoming a key issue in Information Extraction, Question Answering, Summarization, and, in general, in all NLP applications requiring some kind of semantic interpretation [1].

Extracting thematic roles (also called semantic role labeling -SRL) involves identifying which groups of words (phrases) act as the arguments to a given predicate. These arguments must be labeled with their role with respect to the predicate, indicating how the proposition should be semantically interpreted.

Most of the recent works on role labeling exploit statistical methods which depend on supervised learning over statistical features extracted from a semantically-labeled (often manually-created) corpus. These approaches applies a single learning method such as pure statistical models[2], maximum entropies[3], support vector machine[4] and generative models[5] or a combination of them[2] to recognize semantic arguments of verbs. All of these methods need large, semantically labeled corpora. On the other hand there are symbolic methods which need large amounts of lexical and ontological knowledge to assign thematic roles to syntactic constituents. These methods are suitable for cases in which there is no tagged corpus available (as in Persian).

On the other hand most existing systems for automatic semantic role labeling make use of a full syntactic parse of the sentence in order to define argument boundaries and to extract semantic relations between arguments. As complete parsing of sentences have a high cost, especially for languages (such as Persian) for which there is no complete computational grammar, using shallow parsing is a good alternative.

In this paper we propose an approach to assign semantic (thematic) roles to syntactic constituents of a Persian sentence.

There is one another work which implicitly assigns semantic roles for Persian texts. This work has been done in Hasti ontology learning system [6]. Hasti uses full parse trees of simple Persian sentences and assign limited thematic roles (no conflicts, no ambiguity) to constituents by exploiting some rules and simple heuristics. The work done in Hasti regarding thematic role assignment is applicable for restricted circumstances.

Our proposed SRL system implements a two-phase architecture to first identify the arguments and then to label them for each predicate.

As Persian is almost a free word order language and this property results in high structural ambiguity, applying a shallow parsing method can make significant improvements in argument identification. Although the recent researches on shallow parsing have focused on the statistical methods but due to the lack of the suitable linguistic resources in Persian such as tagged or annotated corpus, the rule-based method is exploited and the rules are designed manually. The method proposed to identify the arguments uses the constituent ordering of phrases to determine boundary and type of each phrase.

The proposed system takes advantage of predicate-argument structures. After determining the boundaries of phrases, a set of rules are defined to identify the semantic role of each phrase based on the syntactic and semantic properties of both phrase and the associated predicate.

The rest of this paper is organized as follow: section 2 briefly describes thematic roles and discusses our proposed role set. Section 3 introduces the general architecture of our model and describes its components in details. The experimental results are shown in section 4. Finally, conclusion of this study is presented in section 5.

In all examples throughout this paper, we will show Persian sdentences by their transliteration in italic between quotes

Manuscript received September 9, 2007.

Mehrnoush Shamsfard is an assistant professor at Electrical and Computer Engineering Department, Shahid Beheshti University, Tehran, Iran. (corresponding author to provide phone: +98-21-29902283; e-mail: m-shams@sbu.ac.ir).

Maryam Sadr Mousavi is a researcher at NLP lab, Shahid Beheshti University and also M.Sc. Student at Azad University, Qazvin, Iran (email: maryam_sadrmousavi@yahoo.com).

followed by their translation to English between parentheses.

II. THEMATIC ROLES

Thematic roles, also called thematic relations, semantic roles or θ-roles, are characterizations of certain semantic relationships which hold between a verb and its complements (and adjuncts). For example in the following sentence :

'*Ali ketabha ra az london beh tehran ferestad*.' (Ali sent the books from London to Tehran.)

'Ali' is the Agent, '*ketabha*' (the books) is the Theme or Patient, 'London' is the Source, and 'Tehran' is the Goal or Destination of the sending event denoted by the sentence. Semantic roles are one of the oldest issues in linguistic theory that were first mentioned by Jeffrey Gruber [7]. There is no standard set of semantic roles, nor about their nature or their status in linguistic theory. The set of roles proposed by linguists range from very specific to very general [3]. At the specific end of this spectrum are domain-specific roles applied in some information extraction systems such as the FROM-CITY, TO-CITY, or RECEIVE-TIME roles, which can be applied in reservation systems, or verb-specific roles such as BUYER, GOODS and SELLER for the verb buy. The other end of the spectrum consists of theories with only two "proto-roles": PROTO-AGENT and PROTO-PATIENT [8]. In between there are many theories which propose the limited number of roles (approximately ten roles), such as Fillmore (1971)'s list of nine: AGENT, EXPERIENCER, INSTRUMENT, OBJECT, SOURCE, GOAL, LOCATION, TIME and PATH.

For the task of this paper, we initially employed the role set proposed by Fillmore and then a number of roles are added to provide more abstract semantic characterization.

Our proposed role set consists of 16 roles which are divided into two classes: primary and general roles. The primary roles are the roles which are predicate-specific such as Agent, Amount, Patient, Theme, Instrument, Beneficiary, Cause, Force, Experiencer, Goal and Source. For different predicates some subset of these roles may be available. The second class of roles which are called general are those which are assumed to apply across all verbs, including Location, Time, Manner, Reason and Discourse. For example in the sentence '*banabarin ali dirouz bekhater e sarmakhordegi beh madreseh naraft*' (so Ali didn't go to school yesterday because of catching cold) we have the following primary and general roles:

PHRASE	ROLE	ROLE -CLASS
'Banabarin' (So)	discourse	general
'Ali' (Ali)	agent	primary
'*Dirouz*' (yesterday)	time	general
'*Bekhater e sarmakhordegi*' (because of catching cold)	reason	general
'*Madreseh*' (school)	goal	primary

In the next section we will describe our proposed SRL approach.

III. THE PROPOSED APPROACH

Figure 1 shows the overall architecture of our model.

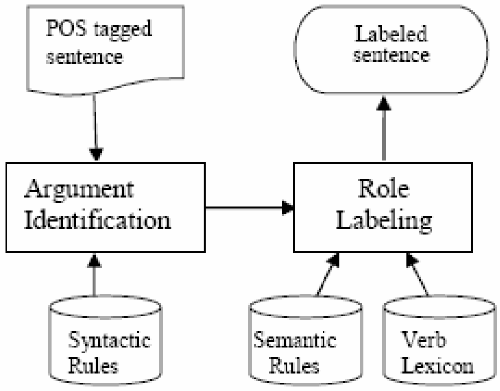

Fig. 1 The Thematic Role extraction system architecture

As it can be seen from the figure, the task of automatic semantic role assignment is divided into two main subtasks: (1) Identification of the target argument boundaries and (2) labeling the arguments with appropriate semantic roles. In other words, given a sentence in natural language, all the predicates associated with its verb(s) have to be identified along with their arguments.

The first part (subtask) can be accomplished by finding all constituents of a sentence through syntactic Rules and checking their POS's.

The second part (subtask) uses some semantic rules to distinguish different roles such as Agent, Goal, etc and also a repository of various Persian verbs and their features. This part faces a complicated problem since the number of arguments and their positions vary depending on a verb's voice (active/passive) and sense, along with many other factors.

The exploited approach to perform both the detection and the classification of predicate arguments is summarized by the following steps:

1. Given a source sentence, perform some morphological and lexical analysis,
2. Execute the shallow syntactic parsing module to recognize the boundaries of arguments,
3. Find the target verb of the sentence and search for it in the verb lexicon to determine its corresponding verb class,
4. Execute semantic role labeling module to assign appropriate semantic role to each argument

In the rest of this section we will provide more detailed information about the two phases.

A. Argument Identification

The identification process is related to determining the constituents' boundaries in the sentence. These constituents represent semantic arguments of a given predicate (often shown by the verb).

Many existing SRL systems usually use a pure syntactic parser to retrieve possible constituents. Once the boundary of a constituent is defined, there is no way to change it in later phases.

In this paper, we have followed X-bar theory to select and identify syntactic chunks or phrases such as noun phrases, prepositional phrases and adverb phrases. In this theory, four syntactic groups are presented through the specific nested structure as a combination of headword, adjunct constituents and determiners [9].

To identify phrase boundaries, first, the common constituent ordering patterns for Persian language are extracted based on this theory. Then a set of rules are introduced to identify the position of each word in its corresponding phrase structure regardless of the phrase type. To discover these positions, we exploited the POS tags assigned to every token by the POS tagger. The chunk boundaries are identified by these handcrafted linguistic rules, which check whether two neighboring POS tags belong to the same chunk, or not.

The next step is to assign a type to each bounded phrase. We defined some rules to determine the phrase type according to the POS tags of tokens (chunk or phrase elements). For example, a chunk that begins with a token with POS tag P will be tagged as PP.

A Problem may occur when there is a mismatch (difference) between syntactic type and semantic type of a phrase. To solve this problem, we added some additional rules to assign some priority to (often) semantic types. As an instance we introduced some rules to recognize the NPs or PPs, which have adverbial meaning. For example the phrase beh sora't (quickly, rapidly) is syntactically a prepositional phrase while it always plays a role of adverb in Persian and so semantically is an adverb phrase. Note that this system is also able to assign both of the mentioned types to such phrases. Adding this property to the system allows users to make use of its results for semantic researches on the language.

1) Tagging style

Chunks information in a sentence can be represented by means of tags. The bracket style and IOB tag set are the two common tagging styles. Bracket style is the simplest case in which the start and end of phrases are limited with brackets. The following sentence is marked using brackets.

[*in ketab* NP] [*bist safheh* NP] [*darad* VP].
[This book NP] [has VP] [twenty pages NP].

In this paper, the alternative style for representing chunks is IOB form [10]. In this scheme, each token is tagged with one of three special chunk tags, I (inside), O (outside), or B (begin). A token is tagged as B if it marks the beginning of a chunk. Subsequent tokens within the chunk are tagged I. All other tokens are tagged O. The B and I tags are suffixed with the chunk type, e.g. B-NP, I-NP. One advantage of encoding chunk structure with tags corresponded to words over bracket markers is that it is not encountered the problem of dependency of words within a phrase. Since this scheme can be easily used in files and also the resulted outputs can be efficiently applied in the different machine learning techniques, using this tag set is proposed. The sentence of the above example can be represented using this notation as in figure2.

in	this	B-NP
ketab	book	I-NP
bist	twenty	B-NP
safheh	page	I-NP
darad	has	B-VP
.	.	O

Fig. 2 IOB tagging for the sentence "this book has twenty pages" (IOB tags are suffixed with the chunk type)

2) Constituent ordering and rule extraction

To identify phrase boundaries we introduced about 60 patterns (rules) for constituent ordering of Phrases in Persian, as well as a description of their structure. These patterns Show that how the lexical information present in the sentence could be used in determining the boundaries of the phrases. The extracted rules can be divided into two groups. The first are the rules that determine if the token is inside the phrase and the next are those that determine the beginning token of a phrase regardless of the phrase type. In other word, these rules can be considered as two classes, the rules mark the token as I or the rules mark it as B.

In the following, we will introduce a sample rule, which is extracted from PP's constituent ordering.

For example, Persian prepositional phrases, however, are easily recognized and can be used to mark phrasal boundaries in the sentence. The following structure describes the constituent ordering of PP.

PP   preposition + NP

The headword of a PP is a preposition, which is always followed by NP. This structure shows that detecting start of a PP is not difficult. The following rule shows that if the current token is a preposition the next token certainly will be in the PP structure.

IF POS (X) = P then IOB-tag (X+1) = I

But the preposition itself is not always the beginner of the phrase. There may be an identifier before it in the prepositional phrase (e.g. '*hatta dorost dar khiaban*' (even right in the street)). Following is a sample rule to handle such cases

IF POS(X) =P and X-1 □generalID then IOB-tag(X) =I

In some cases with ambiguities, assigning the IOB tag is not easy. For example in cases which a preposition occurs in an NP not a PP we may have some ambiguities. As an example in sentence: '*nameh-ye Ali beh Hassan resid*.' (Ali's letter for Hassan, arrived) or (Ali's letter arrived to Hassan) we have two interpretations. For such cases we have developed some disambiguation modules which find the correct chunking regarding statistical and the semantic information about the verb (coded in the verb lexicon) and the constituent. For instance if the verb was '*gom shod*' (was lost) instead of '*resid*' (arrived) in the above example, as the verb – was lost – accepts one argument and has no prepositional phrase with '*beh*' (to), we could choose the first interpretation (considering the preposition inside a noun phrase) easily.

On the other hand we used the statistical information about the probability of initiating an argument (a role) by a preposition (e.g. using preposition '*beh*' (to) for denoting destination role) to disambiguate the chunking process too.

Some of other extracted rules are as following. A detailed description of extracted rules can be found in [14].

IF POS (X) = NUM & POS (X-1)=UNT then IOB-tag (X)= I
IF POS(X) = P then IOB-tag (X+1) = I
IF W(X) = '*ra*' then IOB-tag (X) = I & IOB-tag (X+1) = B
IF POS(X) =SADJ & POS (X+1) =N then IOB-tag (X+1)= I
…

B. Argument Labeling

After identifying the arguments, it's time to tag them with semantic roles. This task is done exploiting a verb lexicon and a set of rules.

1) The Verb Lexicon

The verb lexicon contains the information about verbs (predicate) such as their type (class) and their arguments. Verb classes have been proven to be useful in various (multilingual) natural language processing (NLP) tasks and applications, such as computational lexicography, language generation, machine translation and word sense disambiguation. The main practical aim of verb semantic classifications is to organize them in a verb lexicon according to their common semantic features. Fundamentally, such classes define the mapping from surface realization of arguments to predicate-argument structure. The relationship between such surface manifestations and semantic roles is the subject of linking theory. This theory argues that the syntactic realization of arguments of a predicate is predictable from semantics. This relationship indicates that it may be possible to recognize semantic relationships from syntactic cues.

Although several classifications are now available for English verbs [11, 12], there is no such classification for Persian verbs. In this work, we provided a classification for Persian verbs consist of 22 classes which groups on the basis of both syntactic and semantic alternations. For this purpose, we first grouped a number of Persian verbs (more than 300 verbs at the first stage) according to the number of syntactic arguments and then classified them into smaller groups which have similar set of semantic roles. Each verb will belong to a class with a specific class number and each class can have one to four arguments, with the majority of them having one and two arguments. For example verb khord (eat) belongs to verb class 12 which is described as follow:

Verb class 12:

[+ Agent
+ Theme
Instrument]

This sample representation describes that this verb class must have both agent and theme roles. However, the instrument role is optional.

Table 1 shows the features of 22 proposed verb classes:

TABLE I VERB CLASSES WITH THEIR FEATURES

Class No.	Features
1	[+agent, companion]
2	[+force]
3	[+experiencer]
4	[+(agent or force)]
5	[+agent, source, destination, companion, instrument]
6	[+agent, +patient, instrument]
7	[+agent, +topic, instrument]
8	[+(agent or patient), (topic or instrument)]
9	[+ agent + theme, instrument, companion]
10	[+agent, +theme, instrument, source, destination]
11	[+agent, +patient, beneficiary, instrument]
12	[+agent, +theme-1, theme-2, instrument
13	[+agent, +theme, beneficiary, instrument]
14	[+experiencer, source, destination]
15	[+ agent, +topic]
16	[+ agent, +message]
17	[+agent-1, +agent-2]
18	[+agent, +theme, +beneficiary, instrument]
19	[+ (agent or force), +theme, instrument] or [+ experiencer, (instrument or agent)]
20	[+agent, +patient, +(topic or message)]
21	[+(agent or force or experiencer or cause) +theme, +destination]
22	[+agent, + theme, +destination]

The problem, which has to be solved here, is that a certain set of roles must be associated with a verb class and null instances have not been considered. For example although the roles agent and theme are essential for verb "to eat", in some cases they are omitted in the sentence. Our system detects such instances as semantically ambiguous cases and some works have been done to solve this problem.

2) The Semantic Rule Base

As shown in figure 1 the labeling part of the system consists of a set of handcrafted rules. To extract these rules we have used the following attributes from the results of the shallow parser and the verb lexicon:

- Phrase Type – The syntactic category (NP, PP, SP, etc.) of the phrase corresponding to the semantic role.
- Verb class - the corresponding class of the active predicate described in the verb lexicon
- Head Word - The syntactic head of the argument constituent
- Voice – This attribute distinguishes between active and passive verb.

The rules can be divided into two groups:

1. Declaring Semantic Roles. Semantic roles can be declared based on syntactic and lexical information about the argument. In addition to our heuristics to identify the semantic roles we have also applied the results of the study about

semantic representation of Persian prepositions [13]. For example, we can use any of the following sets of properties to determine the Destination role:

a.
Phrase type: PP
POS of headword: N
Semantic lexical information of head word: +LOC
The beginning word of the phrase: 'beh' (to)
Example: 'beh madrese e ma' (to our school)

b.
Phrase type: PP
POS of headword: N
Semantic lexical information of head word: -ALIVE
The beginning word of the phrase: 'rouy' (on, above), 'zir' (under), 'bala' (up), 'paiin' (down)'
Example: 'rouy e miz'(on the table)

2 Assigning Semantic Roles. These rules specify how to map each argument to one semantic role, based on the class associated to the target verb of the sentence.

The distinction between active and passive verbs plays an important role in the assignment process. Since direct object of active verbs often corresponds to subjects of passive verbs different verb classes have been assigned to the verb's passive form.

To resolve cases in which more than one rule fires, we have defined some weighting factors for the rules based on thematic hierarchy. This approach assumes an ordered list of semantic roles [11] as well as an ordered list of syntactic relations.

IV. Experimental results

The proposed system is implemented and tested. To show sample outputs of the system, we show the results of shallow parsing and role assignment for sentence: '*Ali kheili khoshhal ba pedarash ba mashin beh madreseh raft ta dars bekhanad*' (Ali very happily went to school with his father by car to study) in figure 3.

In order to evaluate the role labeling system, two types of experiments were performed. The first type presents the correctness of the argument identification phase (shallow parsing) and the second type shows the role labeling accuracy. Since the correct assignment of semantic roles is closely related to the correct identification of all arguments in a sentence, we calculated two measures to show the performance of the shallow parser. The first measure presents the percent of correctly detected phrases while another measure is related to the percent of sentences which all of its phrases are detected correctly. So we tested the shallow parser on different sets of randomly selected sentences from texts with different topics, which were hand labeled with POS tags. The results for argument identification phase of the system on 300 randomly selected sentences are shown in Table 2-3.

'*Ali*' (Ali)	B
'*kheili*' (very)	B
'*khoshhal*' (happy)	I
'*ba*' (with)	B
'*pedarash*' (his father)	I
'*ba*' (by)	B
'*mashin*' (car)	I
'*beh*' (to)	B
'*madreseh*' (school)	I
'*raft*' (went)	VB
'*ta*' (to)	B-SP
'*dars*' (lesson)	I-SP
'*bekhanad*' (study)	I-SP
.	O

(a)

[Ali NP] ['kheili khoshhal' (happily) ADVB] ['*ba pedarash*' (with his father) PP] ['*ba mashin* ' (by car) PP] ['*be madreseh*' (to school) PP] ['*raft*' (went) V] ['*ta dars bekhanad*' (to study) SP]

(b)

'*Ali*' (Ali)	Agent
'*pedar*' (father)	companion
'*khoshhal*' (happy)	State
'*mashin*' (car)	Instrument
'*madreseh*' (school)	Destination
'*dars bekhanad*' (study)	Reason

(c)

Fig. 3 results of processing the sample sentence: '*Ali ba pedarash ba khoshhali ba mashin beh madreseh raft ta dars bekhanad*' (Ali went to school happily with his father by car to study)
(a) IOB tags (b) Shallow parsing (chunking) results
(c) role assignment

TABLE II EXPERIMENTAL RESULTS OF THE SHALLOW PARSER

Correctly detected phrases	83%
Correctly detected sentences	85%

TABLE III PRECISION OF THE PARSER FOR EACH PHRASE TYPE

Phrase typet	Percentage of Correct detection
NP	%76
PP	%79
ADJ	%84
ADVP	%86
SP	%90

The second type of experiments also have been performed on the same sentences with this exception that the sentences with the finite (copula) predicates are eliminated (this type of predicates is not mentioned in the verb lexicon). The final results of the role labeling phase are shown in tables 4-5.

Table 4 shows the precision and recall measures for the thematic role extractor, supposing that its input is correct (ignoring the errors caused by the parser) and table 5 shows the overall precision and recall of the system. These measures are calculated for each thematic role separately.

TABLE IV EVALUATING THE THEMATIC ROLE ASSIGNER

Thematic role	Precision (%)	Recall (%)
Agent	89	82
Experiencer	80	75
Patient	84	79
Theme	84	78
Time	86	80
Location	87	80
Source	97	90
Destination	96	90
Reason	89	81
Topic	88	80
Instrument	91	86
Force	89	81
State	77	69
Companion	84	79
Message	90	85
Beneficiary	82	83
Cause	87	80

V. Conclusion

In this paper, a Persian semantic role labeling system based on a set of handcrafted rules is proposed. In general, these rules can be divided into two groups: the rules related to the identification of constituents and the rules for semantic labeling of these constituents. Applying a rule based method has several advantages, the foremost of which is that it eliminates the need to a role labeled corpus, a very expensive resource to produce. In addition the verb semantic lexicon created in this work can be applied in various other natural language processing tasks. Other key advantages of the current system are (1) it labels each argument in a sentence dependent of the others and (2) using shallow parsing to find phrase boundaries eliminates the problems encountered in deep syntactic parsing. Some of the drawbacks of the system which are going to be fixed are (1) it does not consider the copula (finite) verbs and (2) it needs that the Ezafe sign be present implicitly.

TABLE V EVALUATING THE TOTAL SYSTEM

Thematic role	Precision (%)	Recall (%)
Agent	83	76
Experiencer	76	70
Patient	77	72
Theme	77	71
Time	79	73
Location	80	73
Source	91	83
Destination	90	82
Reason	82	75
Topic	81	73
Instrument	87	82
Force	82	75
State	71	66
Companion	79	71
Message	86	81
Beneficiary	87	81
Cause	80	73

References

[1] X. Carreras and L. Màrquez. "Introduction to the CoNLL-2004 shared task: Semantic role labeling". *In Proc. of CoNLL-2004,* Boston, MA, 2004.
[2] D. Gildea, D. Jurafsky: "Automatic labeling of semantic roles", *Computational Linguistics*, 28(3), 2002, pp. 245–288.
[3] J. H. Lim, Y.S. Hwang, S.Y. Park, and H.C. Rim. "Semantic role labeling using maximum entropy model". *In Proc. of CoNLL- 2004*, 2004.
[4] S. Pradhan, K. Hacioglu, V. Krugler, W. Ward, J. Martin, D. Jurafsky, "Support Vector Learning for Semantic Argument Classification", *Machine Learning journal*, 60(1), 2005.
[5] C. A. Thompson, R. Levy, C. D. Manning. "A generative model for semantic role labeling", *In Proc. of ECML'03*, Dubrovnik, Croatia. 2003.
[6] M. Shamsfard, A. A. Barforoush, "Learning Ontologies from text", *International Journal of Human-Computer Studies*, Vol. 60, Jan 2004, pp.17-63.
[7] A. Wagner, "Learning Thematic Rrole relations for lexical semantic nets", PHD thesis, Tubingen university, 2004.
[8] D. Dowty, "Thematic Proto-roles and Argument Selection". *Language* 67, 1991, pp..547–619.
[9] M. Meshkatodini. "How adjuncts add to Persian phrases", *In proc. Of the second workshop on Persian language and computer.* , 2006, pp. 25-41.
[10] E. Tjong, K. Sang, J. Veenstra, "Rrepresenting text chunks", I*n Proc. of EACL'99: Ninth Conference of the European Chapter of the ACL,* Bergen, Norway, 1999,pp. 173–179
[11] R. Jackendoff,, *Semantic Structures*, MIT Press, Cambridge, MA, 1990.
[12] B. Levin, "Semantic Prominence and Argument Realization", *In Proc. Of 74th Annual LSA Meeting*, Chicago, 2000
[13] N. Mansoory, Z Abolhasani, M. Ghayumi. "Semantic representing of Persian prepositions" , *In Proc. Of the first workshop on Persian language and computer*. 2004, pp. 105-111.
[14] M. Sadrmousavi, Shamsfard, M., "A Rule-based Semantic Role Labeling Approach for Persian Sentences", *In Proc. Of The Second Workshop on Computational Approaches to Arabic Script Languages (CAASL2)*, Stanford, USA. 2006.

Mehrnoush Shamsfard received her B.Sc in computer software engineering in 1991 and M.Sc in the same field in 1995 both from Sharif university of technology, Tehran, Iran. She received her Ph.D. in Computer engineering-artificial intelligence from Amirkabir university of technology in Jan. 2003.
She is now an assistant professor at electrical and computer Engineering department, Shahid Beheshti University (SBU), Tehran, Iran. She is the founder of NLP Research Laboratory at SBU and her research works are mainly focused on natural language processing and ontology engineering. She has several papers on these areas and her book on ontology engineering is to be published soon.

Maryam sadr Mousavi received her B.Sc in computer software engineering in 2004 from Azad University, Qazvin, Iran. She is currently a graduate student in M.Sc. program in the same university. Her M.Sc. thesis is about Extracting thematic roles from Persian sentences.