Information Structure in Persian: A Comparison of Systemic Functional Grammar & Role and Reference Grammar

Setareh Majidi
Department of Linguistics, Science and Research Branch, Islamic Azad University, Tehran, Iran
Email: majidi_setareh@yahoo.com

Mohammad Dabirmoghaddam
Department of Linguistics, Allameh Tabataba'i University, Tehran, Iran
Email: dadirmoghaddam@yahoo.com

Abstract—The aim of this paper is to compare the approaches to Persian information structure in two functional theories: Systemic Functional Grammar & Role and Reference Grammar. By selecting 400 data from scientific and educational texts, stories and newspapers we have analyzed the relationship between information structure in one hand, and phonology, semantics, morphology and syntax in another hand based on Halliday's Systemic Functional Grammar & VanValin's Role and Reference Grammar. Also, advantages and disadvantages of the mentioned approaches in relation to information structure are investigated. After discussing what is meant by information distribution in each theory, we examine the classification of focus in each approach. Finally the obtained results of Persian data show that the relationship between information structure in one hand and phonology, semantics, morphology and syntax in another hand is not regular and rule-governed. Moreover, studies represent despite the different analytical methods and the shortcomings of each theory, the obtained outcomes in the framework of these two approaches have some in common because of similarity in functional nature of the both.

Index Terms—information structure, Systemic Functional Grammar, Role and Reference Grammar

I. INTRODUCTION

The study of information structure goes back to the beginnings of modern linguistics, to the work of the Czech linguist Mathesius in the 1920s. In recent years advances in understanding how information structure affects syntactic structure have been made by Firbas (1964,1974), Halliday (1967,1985), Kuno (1972), Chafe (1976,1987), Prince (1981,1992), Givon (1984), Lambrecht (1986,1987,1994), Vallduvi (1990,1992), Dryer (1996), VanValin & Lapolla (1997), and others. Nowadays, information structure is one of the crucial subjects in linguistics. In spite of accepting such structure by linguists, they do not have common ideas about the description of this structure and the determination of its categories. The two theories under scrutiny in the present paper Systemic Functional Grammar & Role and Reference Grammar, all make use of the concept of information structure. The continuity of such researches will provide researchers, trainers, and practitioners with valuable findings in comparing the approaches to information structure in two theories, more understanding the nature and application of information structure in Persian, codifying the comprehensive Persian grammar, decoding the complications of discourse comprehension in Persian, and making the opportunity for other discourse-based researches in Persian. The findings of this paper can be helpful for any research such as (translation, discourse analysis, Persian teaching to non-native speakers …). We shall first look at what is meant by information distribution and classification of focus in each theory. We shall find that in general SFG and RRG have some in common. We then examine data which highlight the similarities and differences, in order to assess the advantages and disadvantages of different approaches. In this research, by selecting 400 data from scientific and educational texts, stories and newspapers we have analyzed the relationship between information structure in one hand, and phonology, semantics, morphology and syntax in another hand. Information structure in both theories of SFG and RRG involves so much unsaid. The authors believe that using the outcomes of each approach can be helpful in completion of shortcomings of each one. The importance of information structure is in making correlation among syntax, semantics and pragmatics. SFG and RRG as a universal functional theories are able to analyze and justify Persian grammatical subjects from new viewpoint.

II. WHAT IS MEANT BY INFORMATION DISTRIBUTION?

A. Information Distribution in SFG

The approach to information structuring taken in SFG derives from Prague school ideas, but with one particularly

important difference. Whereas Mathesius writes of Theme as …that which is known or at least obvious in the given situation and from which the speaker proceeds (translated in Firbas, 1964, p. 268).

Halliday separates out two strands of meaning, each corresponding to one half of the Mathesius definition: 'that which is known or at least obvious in a given situation' is given information (information that is presented by the speaker as recoverable […] to the listener' (Halliday and Matthiessen, 2004, p. 91) rather than new information (presented as non-recoverable), while that 'from which the speaker proceeds' is what is regarded as the Theme in SFG. Thus, the given / new distinction is listener-oriented, while the Theme / Rheme distinction is speaker-oriented (Halliday and matthiesen, 2004, p. 93).

Information focus, as defined in SFG, is concerned only with the given / new distinction encoded in prosodic phonology. An information unit can consist wholly of new information, though more usually there is some given information to act as an established background against which the new material is presented. Thus, the information unit is made up of an obligatory element with the function New and an optional element with the function Given (Halliday and Matthiessen, 2004, p. 89). The element in the tone group bearing the main pitch movement (tonic prominence) is seen as carrying information focus, characterised by Halliday as follows:

Information focus reflects the speaker's decision as to where the main burden of the message lies. It is one of the many diverse phenomena referred to by speakers of English as 'emphasis'; […] information focus is one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block as that which he wishes to be interpreted as informative. What is focal is 'new' information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse (Halliday, 1967, p. 204).

B. Information Distribution in RRG

The account of information structuring given in RRG (Van Valin & Lapolla, 1997) is based on the work of Lambrecht, who makes a distinction which has important consequences for the treatment of information distribution:

… a distinction between (i) the pragmatic states of the denotata of individual sentence constituents in the minds of the speech participants, and (ii) the pragmatic relations established between these referents and the propositions in which they play the role of predicates or arguments. It is the establishment of such pragmatic relations that makes information possible (Lambrecht, 1996, p. 49).

The 'old' information is the set of assumptions evoked by the understanding the utterance. We will now refer to this set of assumptions as the 'pragmatic presupposition' or just 'presupposition'. The part of the assertion which is not within the pragmatic presupposition we will call the 'focus' or 'focus of the assertion'; it is the part that is unpredictable or unrecoverable from the context (Van Valin & Lapolla, 1997, p. 202).

III. CLASSIFICATION OF FOCUS

A. Classification of Focus in SFG

In SFG the only classification of focus presented is between marked and unmarked information focus. SFG would classify the information focus in each tone group as unmarked, since it occurs on the last content item. If there is a focused non-deictic adjunct in final position in a tone group, SFG will treat this as an example of unmarked information focus. Since the intonational prominence is not on the last lexical word in the tone group, so that SFG must treat it as a marked information focus.

B. Classification of Focus in RRG

The classification of focus in RRG (Van Valin & Lapolla, 1997, PP. 206-210) is as shown in Figure 1.

Predicate focus represents the unmarked type and corresponds to the type of structure often referred to as topic-comment. Sentence focus is as a case of an 'all new' predication, there being no topic and indeed no pragmatic presupposition, the point of the sentence being to present a new situation. Narrow focus picks out a single clause constituent. According to Van Valin and Lapolla (1997, p. 209), narrow focus can be unmarked or marked, depending on its position with respect to the syntactic structure. In RRG, the syntactic structure of the clause is represented in terms of core and periphery. The core consists of the nucleus, containing the predicate, while periphery contains adjuncts and any non-central arguments. The unmarked position for focus in English is the last element of the core, which of course need not be the last element of the clause, if there are also post-core peripheral elements.

(1) a) Mâšinat chi šode?
    b) Mâšinam **xarâb šode**.   (Predicate focus)
(2) a) Chi šode?
    b) **Mâšinam xarâb šode.**   (Sentence focus)
(3) a) Mâšinat xarâb šode?
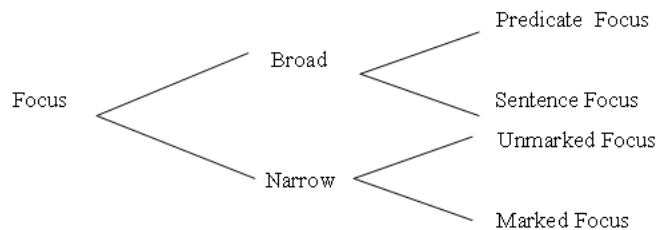   b) Na, **motoram** xarâb šode.   (Narrow focus)

Figure 1. Classification of focus in RRG.

IV. DATA ANALYSIS IN SFG

1) Data analysis based on SFG shows that information structure is determined by the thematic structure. That is to say that, information status of theme/rheme is evaluated with respect to each other. For example, it can be a clause which its theme is older/newer than its rheme and visa versa. Therefore, having old/new information is a relative and gradable concept. Since the mapping of theme/rheme structure on old/new information in the Persian sentence has a high frequency, therefore, such structure is unmarked, whereas the mapping of theme/rheme on old/old information has a low frequency and this is a marked structure.
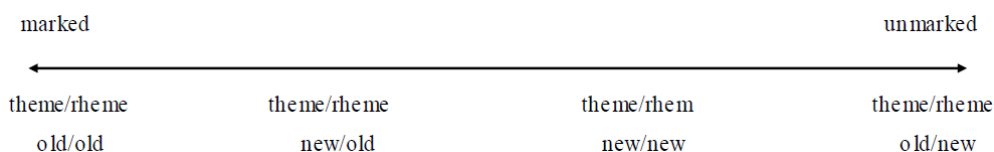
2) Statistical results (Table I) show that 72.33% of the theme/rheme structures have old/new information and 21.72% of them contain new/new information. Moreover, 4.03% of the theme/rheme structures have new/old information and just 1.92% of them have old/old information.

TABLE I.
FREQUENCY OF INFORMATION STRUCTURE IN THE THEME/RHEME DOMAIN

Thematic structure / Information structure	Theme/Rheme	Percentage (%)
old/new	413	72.33
new/new	124	21.72
new/old	23	4.03
old/old	11	1.92

3) In the theme and rheme domains, verbs contain focal stress with 37.08%, direct objects 23.84%, adverbs and adjuncts 16.78%, Wh–questions 13.69%, indirect objects 6.84% and subjects 1.77% (Figure 2).
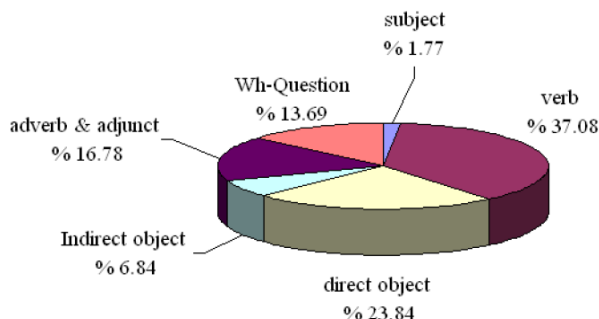
Figure 2 . Frequency of focal stress in the theme/rheme structure

4) In the theme domain with the old and new information, there are arguments (internal and external) with high frequency and there is not any predicate in this domain. In the rheme domain with the old information, there are predicates with 53.49% and arguments with 46.51%. In the rheme domain with the new information, the frequencies of predicates and arguments are 70.20% and 29.80% respectively.

5) The frequencies of lexical categories in the old theme domain are as follows: nouns with 88.72%, adjectives

6.29% and adverbs 4.99%. In the new theme domain, nouns contain 65%, adjectives 17.5%, adverbs 12% and Wh-questions 5.5%. In the old rheme domain, there are nouns with 54.24%, verbs 33.90%, adjectives 10.17% and adverbs 1.69%. In the new rheme domain, the frequency of nouns is 40.25%, verbs 38.07%, adjectives 13.20%, adverbs 7.83% and Wh-questions 0.65%.
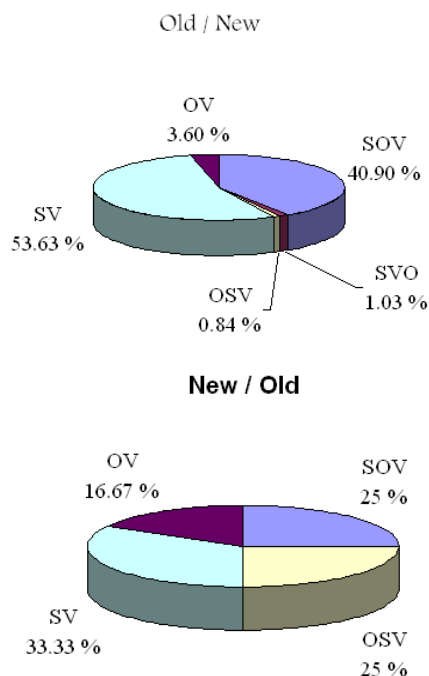
6) In the theme/rheme domains with old/new information, the frequency of (Subject-Verb) is 53.63%, (Subject-Object-verb) 40.90%, (Object-verb) 3.60%, (Subject-verb-Object) 1.03% and (Object-Subject-Verb) 0.84%. In the theme/rheme domains with new/old information, the frequency of (Subject-Verb) is 33.33%, (Subject-Object-Verb) and (Object-Subject-Verb) each one 25%, (Object-Verb) 16.67%. In the theme/rheme domains with new/new information, (Subject-Verb) has 75.97% frequency, (Subject-Object-Verb) 16.27%, (Object-Verb) 5.43%, (Object-Subject-Verb) 1.55% and (Subject-Verb-Object) 0.78%. In the most marked types of information structure that is old/old in the theme/rheme domains, the frequency of (Subject-Verb) is 60 % and (Subject-Object-Verb) 40% (Figure 3).

V. DATA ANALYSIS IN RRG

1) The analysis of data based on RRG shows that determination of information structure in each clause depends on thematic structure of it. It means that the information status of the theme and rheme is evaluated with respect to each other. It can be stated that making presupposition and focus in the theme and rheme are evaluated in comparison with each other; therefore, they are relative and gradable concepts. Thus, there are four types of focus structure for sentences.

2) Also, markedness in relation to information structure is relative and gradable. It is possible to show all kinds of focus structures on markedness continuum. Since predicate-focus structures have 73.10% frequency so they are unmarked. There are Sentence-focus and unmarked narrow-focus structures with 21.95% and 0.88% frequencies respectively. Marked narrow-focus structures have 2.12% frequencies and the frequency of non-focus structures is 1.95% (Table II).

3) The study findings reveal that 30.50 % of the predicate-focus structures are topic-prominent and 69.50% of them are subject-prominent. In sentence-focus structures, the frequency of topic-prominence sentences is 43.55% and subject-prominence is 56.45%. In unmarked narrow-focus structures, the frequency of topic-prominent is 60% and subject-prominent is 40%. In marked narrow-focus structures, the frequencies of topic-prominent and subject-prominent sentences are 33.33% and 66.67% respectively. In non-focus structures, the frequency of topic-prominent sentences is 18.18% and subject-prominent is 81.81%. Wholly, Persian tendency to subject-prominent sentences is more than topic-prominent ones. Thus, 66.55% of the sentences are subject-prominent and 33.45% of them are topic-prominent.
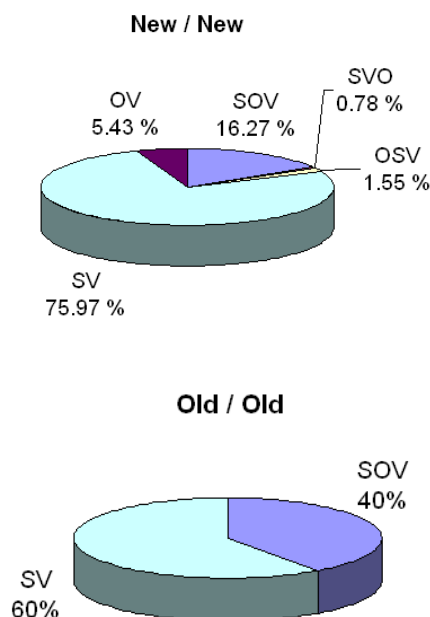
Old / New

OV
3.60 %

SOV
40.90 %

SV
53.63 %

SVO
1.03 %

OSV
0.84 %

New / Old

OV
16.67 %

SOV
25 %

SV
33.33 %

OSV
25 %

New / New

Old / Old

Figure 3 . Frequency of word order in the theme/rheme structure

TABLE II.
FREQUENCY OF FOCUS STRUCTURE

Focus structure	Frequency	Percentage (%)
Predicate-focus	413	73.10
Sentence-focus	124	21.95
Unmarked narrow-focus	5	0.88
Marked narrow-focus	12	2.12
Non-focus	11	1.95
total	565	100

4) Persian is a discourse-configurational language only with respect to topics. Since in Persian, topic is considered the first constituent of the sentence, therefore, it can be stated that topic is specified by word order and Persian is a topic-prominent language.

5) Persian shows that word order is flexible and the constituent structuring is relatively free. Since the entire main clause in Persian can be located in potential focus domain, therefore, it will be considered a discourse-non-configurational language with respect to foci and all the constituents can be focus potentially. Sentence focus depends on word order, stress place and morphology in Persian. The possibility of having focus in all constituents rejects the predictability of stress place. Thus, the place of sentence stress is not regular and rule-governed.

6) The results show that in the predicate-focus structures, the frequency of verbs containing focal stress is 49.01%, direct objects 26.33%, indirect objects 7.09%, adverbs and adjuncts 15.70% and Wh-questions 1.78%. In the sentence-focus domain, verbs contain focal stress with 46.52%, direct and indirect objects respectively 28.4% and 8.53%, adverbs and adjuncts 10.85%, Wh–questions 3.93% and subjects 1.77%. In the unmarked narrow-focus domain, direct and indirect objects contain focal stress with 63.94% and 36.06% respectively. In the marked narrow-focus domain, the frequencies of subjects, adverbs and adjuncts are 21.14% and 78.86%.
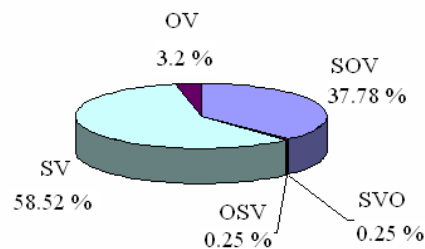
7) In predicate focus structure, the frequencies of predicates and arguments (internal and external) are 70.82% and 29.18% respectively. Whereas in sentence focus structures, the frequencies of predicates and arguments are 48.57% and 51.43%. It can be stated that in unmarked and marked narrow-focus structures, there are arguments with 100% frequency and are not any predicates.

8) To the frequency of lexical categories in the predicate-focus domain, there are nouns with 40.75% frequency, verbs 35.07%, adjectives 14.70%, adverbs 13.89% and Wh-questions 0.8%. In the sentence-focus domain, the frequency of nouns is 33.08%, verbs 32.57%, adjectives 17.43%, adverbs 13.89% and Wh-questions 3.03%. In the unmarked narrow-focus domain, the frequency of nouns is 100%. In the marked narrow-focus domain, we have nouns and adverbs with 21.14% and 78.86% frequencies.
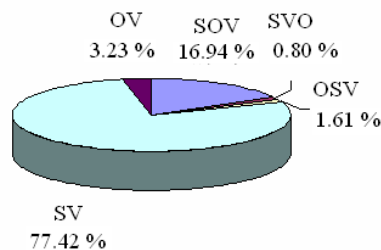
9) In the predicate-focus domain, the frequency of (Subject-Verb) is 58.52%, (Subject-Object-verb) 37.78%, (Object-verb) 3.20%, (Subject-verb-Object) and (Object-Subject-Verb) each one 0.25%. In the sentence-focus, the

frequency of (Subject-Verb) is 77.42%, (Subject-Object-Verb) 16.94%, (Object-Subject-Verb) 1.61%, (Object-Verb) 3.23% and (Subject-Verb-Object) 0.80%. In unmarked narrow-focus domain, (Subject-Object-Verb) has 20% frequency, (Object-Verb) and (Subject-Verb) each one with 40%. In the marked narrow-focus domain, the frequency of (Subject-Verb) is 75%, (Subject-Object-Verb) 8.33% and (Object-Subject-Verb) 16.67% (Figure 4).
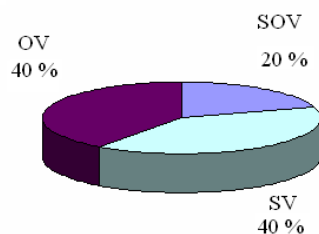
Predicate-Focus

OV
3.2 %

SOV
37.78 %

SV
58.52 %

OSV
0.25 %

SVO
0.25 %

Sentence-focus

OV
3.23 %

SOV
16.94 %

SVO
0.80 %

OSV
1.61 %

SV
77.42 %

Unmarked narrow-focus

SOV
20 %

OV
40 %

SV
40 %

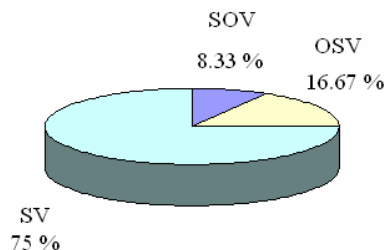**Marked narrow-focus**

SOV
8.33 %

OSV
16.67 %

SV
75 %

Figure 4. Frequency of word order in the focus structure

VI. CONCLUSION

After conducting a statistical analysis, it was found that information structure in both theories of SFG and RRG has so much unsaid. The authors believe that using the outcomes of each approach can be helpful in completion of each one's shortcomings. SFG and RRG as a universal functional theories are able to analyze and justify Persian grammatical matters from new viewpoint. Also, studies represent that relationship between information structure in one hand and phonology, semantics, morphology and syntax in another hand is not regular and rule-governed. Despite the different

analytical method and the shortcomings of each theory, the obtained outcomes in the framework of these two approaches have some in common because of similarity in functional nature of the both. Intonation, word order, morphology and focusing on co-text and context as an independent variables affect on information structure. In Persian, Stress, Morphology, word order or combination of these features are influential factors in markedness of information structure. Persian shows that any phonetic and semantic prominence can be the focus of sentence in this language. Because of scrambling and relatively free word order in this language, each constituent can be located as a sentence focus. Thus, the predictability pattern of stress and sentence focus in Persian faces difficulties. It is so difficult to draw a strict line between configurational / non-configurational languages in one hand and discourse configurational / discourse non-configurational languages in another hand. Therefore, this categorization should be considered hierarchically. In one hand Persian is located in the middle of this continuum syntactically and is considered a relatively non-configurational language, in another hand it is at the end of the continuum on the basis of focus structure and it is a non-configurational language.

APPENDIX A.   SAMPLES OF SFG ANALYSIS

Thematic structure / Sample	Theme	Rheme
1	Agar ø ø	**bexâhim tasviri az chehreye "zabânshenâsiye nazariye" emruz tarsim konim** **mitavânim be jor?at az se negareše mosallat dar in rešte soxan beguyim.**
2	In se negareš	**dar vâqhe se tafakkore qâleb dar zabânšenâsiye nazariye emruz-and,**
3	Be in ta?bir ke har kodâm	**tarafdârân va hâmiyâne qâlebe molâhezeyi râ jazb nemude ast,**
4	**Be tori ke gruh- haye zabânšenâsi, ketâb-hâ, maqâle-hâ, resâle-hâ, majalle-hâ va hamâyesh-hâye elmi**	omdatan be mo?arrefi, paŽuheš va našre ârâ va afkâre naš?at gerefte az in negâreš-ha mipardâzand.
5	In negâreš- hâye mosallat	**ebâratand az: "zabânšenâsiye suratgarâ", "zabânšenâsiye naqšgara", "zabânšenâsiye šenâxti".**
6	Har yek az in negareš-hâ **Ke fasle moštarake ân-hâ**	**zirmajmu?e-hâyi râ šâmel mišavad** **be tartib ebârat ast az talaqiye zabân be onvâne "nezâmi sâxt bonyân va riyâzi gune", "nezâmi barâye ijâde ertebât", va "nezâmi šenâxti.**
7	Be bayâni digar, in se negareš	**dar vâqe? se ruykard be zabân-and**
8	Be in ma?ni ke ø	**az se manzare moxtalef zabân râ ta?rif karde**
9	Va ø	**morede motâle?e qarâr dâde-and.**

APPENDIX B.   SAMPLES OF RRG ANALYSIS

Focus structure / Sample	**Predicate-focus**	*Sentence-focus*	<u>Unmarked narrow-focus</u>	**<u>Marked narrow focus</u>**
1	**Agar bexâhim tasviri az chehreye "zabânšenasiye nazariye" emruz tarsim konim** mitavânim be jor?at az se negareše mosallat dar in rešte soxan beguyim.			
2	In se negareš dar vâqe? <u>se tafakkore qâleb</u> dar zabânšenâsiye nazariye emruz-and,			
3	Be in ta?bir ke har kodâm tarafdârân va hâmiyâne qâlebe molâhezeyi râ jazb nemude ast,			
4	Be tori ke gruh-haye zabânšenâsi, ketâb-hâ, maqâle-hâ, resâle-hâ, majalle-hâ va hamâyesh-hâye elmi omdatan be mo?arrefi, paŽuheš va našre ârâ va afkâre naš?at gerefte az in negareš-hâ mipardâzand.			
5	In negareš- hâye mosallat **ebâratand az: "zabânšenâsiye suratgarâ", "zabânšenâsiye naqšgarâ", "zabânšenâsiye šenâxti".**			
6	Har yek az in negareš-hâ **zirmajmu?e-hâyi râ šâmel mišavad**			
7	Be bayâni digar, in <u>se negareš</u> dar vaqe? se ruykard be zabân-and			
8	Be in ma?ni ke az <u>se manzare moxtalef</u> zabân râ ta?rif karde			
9	**Va morede motâle?e qarâr dâde-and.**			

REFERENCES

[1]   Chafe, W. (1976). "Giveness, Contrastive, Definitness, Subjects, Topics and Point of View In Charles N. Li (ed.), *Subject and Topic.* pp:22-156. New York: Academic Press.

[2]   Chafe,W. (1987). "Cognitive Constraints On Information Flow". Tomlin, R. (ed.), *Coherence and Grounding in Discourse*, pp: 21-51. Amsterdam: John Benjamins.

[3]   Firbas, J. (1964). "On Defining the Theme in Functional Sentence Analysis", *TLP* 1: 267-280.

[4]   Firbas, J. (1974). "Some Aspects of the Czechoslovak Approach to Problems of Functional Sentence Perspective". In F. Danes (ed.), *Papers on Functional Sentence Perspective*, The Hague: Mouton, and Prague: Academic.

[5]   Givon,T. (1984). Syntax: A Functional Typological Introduction, Vol. I & II Amsterdam / Philadelphia: John Benjamins.

[6]   Halliday, M.A.K. (1967). "Notes on Transitivity and Theme in English", *Journal of Linguistics* Vol. 3: 37-81 (Part 1), 199-244 (Part 2).

[7]   Halliday, M.A.K. (1985). An Introduction to Functional Grammar, Baltimor: University Park Press.

[8]   Halliday, M.A.K. & Christian M.I.M. Matthiessen (2004), An Introduction to Functional Grammar, New York: Oxford University Press.

[9]   Kuno, S. (1972). "Functional Sentence Perspective: A Case Study from Japanese and English", *Linguistic Inquiry*, 3: 269-320.

[10]   Lambrecht, K. (1996). Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents. Cambridge University Press.

[11]   Prince, E.F. (1981). "Towards a Taxonomy of Given-New Information". In Peter Cole (ed.), *Radical Pragmatics*. pp:223-255. New York: Academic Press.

[12]   Prince, E.F. (1992). Subject, Definiteness and Information Status. In Discourse Description: Diverse Linguistic Analysis of a Fund-Raising Text, William C. Mann & Sandra A. thompson (eds.), pp: 295-325. Amsterdam / Philadelphia: John Benjamins.

[13]   Vallduvi, E. (1990). "The Information Component", PhD. Dissertation, University of Pennsylvania.

[14]   Vallduvi, E. (1992). The Information Component. New York: Garland.

[15]   Van Valin, Robert D. , Jr. & Randy J. Lapolla (1997). Syntax: Structure, Meaning and Function. Cambridge University Press.

Setareh Majidi was born in Tehran, Iran, in 1981. She received the B.A. and M.A. degrees in English translation and Linguistics from the Azad University, Tehran, Iran, in 2001 and 2005, respectively, and is currently working toward the Ph.D. degree in Linguistics from the science and research branch of Islamic Azad University, Tehran, Iran.

Her research activities and interests include Syntax, Pragmatics, Discourse analysis, and Information Structure. She teaches Linguistics in central branch of Azad University, Tehran, Iran. She has several journal papers. Also, she is a member of Linguistics Society of Iran (LSI).

Mohammad Dabirmoghaddam was born in Kashan, Iran. He received the B.A. and M.A. degrees in Linguistics from Tehran University and the M.A. and Ph.D. degrees from the Illinois at Urbana-Shampaign, U.S.A.

He is currently a professor of Linguistics at Allameh Tabataba'i University, Tehran, Iran. His research activities and interests include Syntax, Information Structure and Language typology. He has authored many journal and conference papers in the field of Linguistics.