



Università Ca' Foscari Venezia

Master's Degree
in Language Sciences

Final Thesis

Study of Thematic role of Persian Complex Predicate

Based on the proto-role
hypothesis

Supervisor

Ch. Prof. Gianluca Lebani

Assistant supervisor

Ch. Prof. Masoud Jasbi (University of California Davis)

Graduand

Aria Rastegar
875555

Academic Year

2020 / 2021

If I have seen further, it is by standing on the shoulders of Giants.

— Isaac Newton

Contents

1	Theoretical Background	5
1.1	Argument Structure	5
1.2	Persian Complex Predicates	7
1.2.1	Approaches towards Light Verbs	9
1.3	Approaches towards the Persian Complex Predicates	11
1.3.1	Lexical approach	12
1.3.2	Constructionist approach	15
1.3.3	Construction-based approach	20
1.4	Thematic Roles	26
1.4.1	Case Roles	27
1.4.2	Lexical Conceptual Structure and Thematic Relation	28
1.4.3	Proto-Roles	30
1.4.4	Semantic Macroroles	32
1.4.5	Thematic roles and Argument Structure Processing	35
1.5	Related Studies	37
1.5.1	Studies on Persian Complex Predicates	38
1.5.2	Studies on Thematic Roles	40
1.6	The Current Study	44
2	Research Methodology	47
2.1	Methods	47
2.1.1	Material	47
2.1.2	Participants	52
2.1.3	Procedure	52
2.2	Data Analysis	54
2.2.1	Data	55
2.2.2	Regression Model	56
3	Results	65
3.1	The role of the Non-Verbal Element	66

3.2	The role of the Light Verb Element	66
3.3	The interaction of the Elements	67
4	Conclusion and Discussion	73
A	NVs and LVs used in the questionnaire	77
B	List of the Questions and the Filler entities	81

List of Figures

1.1	Subtypes of Locatum- <i>zadan</i> Construction	25
1.2	A partial hierarchy of Persian CPrs constructions proposed by Samvelian and Faghiri (2013b)	25
1.3	Continuum from verb-specific semantic roles to grammatical relations	33
1.4	Thematic relations continuum in terms of Logical Structure of argu- ment positions	35
2.1	the frequency of the LVs in UDP corpus	49
2.2	Distribution of Proto-role Entities	52
2.3	screenshot of the questionnaire	54
2.4	Predication of the proto-roles based on the light verbs	61
3.1	The interaction of the non-verbal and proto-roles	67
3.2	The interaction of the light verbs on judgments	68
3.3	The interaction of the non-verbal and proto-roles on the judgements	68
3.4	The interaction of the NV and proto-roles on the judgements of proto-roles	70

List of Tables

1.1	Semantic categorization of Persian Light Verbs proposed by Karimi- dustan and Eshaghi (2019)	12
-----	---	----

1.2	Telicity of the Persian Complex Predicates according to Folli et al. (2005)	19
1.3	Constructions of <i>zadan</i> , 'to hit' proposed by Samvelian and Faghiri (2013b)	24
1.4	Proto-role properties	30
1.5	Lexical representations for Aktionsart classes	32
1.6	Kako's (2005) proto-features	41
1.7	Semantic Proto-role Protocol 2	43
2.1	Selected Light Verbs and their Semantic Classification	48
2.2	An example of the proto-features and their ratings from White et al. (2016)	51
2.3	Proto-role entities used in the questionnaire	53
2.4	Percentage of the Judgments for the proto-roles based on the Light Verb and the Non-Verbal	55
2.5	Comparison of different models, according to the hypotheses	57
2.6	Output of the dummy-coded model	60
2.7	Output of the contrast-coded model	62
3.1	Post-hoc Analysis on the effect of NVs	69
3.2	Post-hoc Analysis on the effect of NVs and LV on proto-role judgements	71
A.1	dirstribution of non-verbals and the light verbs in the questionnaire	79

Abstract

Complex Predicates (CPr)s are one of the most controversial concepts among linguists focused on Iranian languages. One of the topics which have gathered the attention of many scholars considers the thematic role of the constructions and the different components of the complex predicate, namely, the non-verbal element (NV) and the light verb (LV). In this research, we review the literature on Persian complex predicates and discuss their different elements. In addition, we will focus on how these theories, model the role of NVs and LVs in the assignment of the thematic roles.

In continue, we review the approaches toward thematic roles from theoretical aspects and psycholinguistic points of view. Finally, we explain why we have chosen the approach of [Dowty \(1991\)](#) for this research. In this research, by using [White et al. \(2016\)](#)'s dataset, we form a list of entities that have the lexical properties of the proto-roles of Dowty. These entities were used as choices in a questionnaire, which was made of Persian CPr constructions, including constructions with meaningful and nonce NVs.

At the end of this research by deriving empirical data, we propose a model, which can give us a prediction and a trend of how different LVs can assign one of the proto-roles to the subject of the construction, and how the presence or absence of the NV element can affect the reading of the speakers.

The results of the research indicate that an interactive account of the CPr elements on the thematic roles is the optimal model; suggesting that, both of the elements have an impact on the understanding of the proto-roles; however, their effects in all of the instances are not the same.

Acknowledgments

This master thesis is about the agent and the patient in the sentences; it is important to note that if an agent has finished a job successfully, it is thanks to the help and support of their beloved ones. I would like to take this opportunity to thank all of the people who supported and helped me during this journey. Through the years, many people have generously helped me with their knowledge, skills, time, kindness, and support. Words do not do justice to what they have done for me, but in thanking them here, I hope I succeed in expressing my sincere gratitude. I will never forget the influence they have had on my development, and I will always strive to pay their kindness and generosity forward.

I would like to thank the linguistics department of the Ca' Foscari University of Venice, who supported me and introduced me to the world of linguistics and supported during these years. I would like to particularly express my gratitude to my supervisor, Dr. Gianluca Lebani, who patiently listened, supported, assisted, and helped me in the course of this research. I would like to express my thankfulness to my co-supervisor, Dr. Masoud Jasbi, who generously dedicated many hours, to talking to me and guiding me on this route.

I would like to also thank my supportive friends. Friends who motivated me and supported me during this MA program.

I am also eternally grateful for the sacrifices that my parents have made in the course of my education. All I have achieved and all I will achieve truly belong to them. Finally, I would like to thank Reihaneh: for being who she is and for who she has helped me to be.

Venice, June 2022

Aria R.

Introduction

In the literature on thematic roles, the story goes by that these linguistic roles are said to be one of the oldest notions in linguistics and human language. One of the oldest references goes back to almost twenty centuries ago in Panini's Sanskrit grammar. Panini mentions several roles for the elements of a sentence and labels them according to their meaning. Although the notion of thematic role is one of the most tangible linguistic notions, linguists do not seem to have a consensus on the exact definition of these roles. These differences in the understanding of the thematic roles have resulted in a relatively large number of frameworks and theoretical debates among linguists; however, perhaps the majority of the scholars agree that these concepts are highly related to the event's structure. In addition, scholars in cognitive sciences and psycholinguistics argue that some of these roles, such as *agent* and *patient* rather than being linguistic notions are part of the core knowledge of human beings (Strickland, 2017).

Among the frameworks dealing with the thematic roles, Dowty (1991)'s proto-role framework assumes these roles are two general and prototypical notions. This is interesting because it is the closest representation of these roles with what our mind uses as a schema to comprehend the verbal arguments of the sentences. The other critical aspect of proto-roles is related to the fact that these roles are decomposed into several lexical and semantic features, enabling researchers to investigate them.

While thematic roles are essential notions in our everyday language understating, the main focus of the research is not on the notion of thematic roles and their underlying theoretical backgrounds directly. In this research, these roles have been investigated and used to answer several other questions regarding their reading in certain structures. Complex Predicates have been among the most studied constructions in Persian linguistic studies. The constituting elements of these constructions, namely, the Non-verbal element and the Light Verbs, have been the center of attention of many scholars regarding their role in syntax and semantics.

The interesting point in studying thematic roles in complex predicates goes back to the fact that there are two constituting elements that are shaping this frame in complex predicates. It is relatively vague in the literature which of these elements

are more effective in allocating the roles to the arguments of the sentences. In the previous studies, different interpretations have been proposed; some scholars believe that the non-verbal elements form the event and select the roles of the arguments; on the other hand, it has been argued that by having a fixed NV and changing the light verb, the roles change. Moreover, some scholars argue that both elements play a significant role, and they need to be studied as a whole and after their incorporation.

This research intends to investigate the character of the constituting elements of the Persian complex predicates in terms of thematic roles based on the empirical data and according to the speakers' judgments. Through this research, we are trying to observe the trends among the contribution of the light verb and the non-verbal elements. The novelty of the present study is that it is not focused on a number of instances for or against the theoretical backgrounds, while it intends to observe and investigate the trends and changes in the effectiveness of the CPr components. The results of the research can find an answer to several questions. Firstly, it will highlight how different light verb elements of the complex predicates can affect the speakers' judgment of the thematic roles. Secondly, it will illustrate the role non-verbal elements play in these judgments. Furthermore, it will characterize the interaction of these elements.

A brief outline of the research is as in the following:

Theoretical Background opens with the discussion on the argument structure and the role of verbs in their formation. Next, I will review the background of Complex Predicates constructions from three approaches of Lexicalist, Constructionalist, and Construction-based perspective, and finally, different frameworks on thematic roles have been described. Next, I will discuss the cognitive science's account of the thematic roles and why we are using the framework of Dowty (1991); ultimately, I will review some of the related studies and discuss the research question

Methodology presents the research methodology, namely, the formation of the questionnaire, the information about the participants, and the procedure. In addition, the results obtained from the research and the methods for the data analysis are described.

Results describes the statistical model and investigates the answer to how the non-verbal and light verb elements can affect the understanding of participants on thematic roles. Moreover, where they interact based on the model.

Conclusion and Discussion , summarizes the study and discusses how this research contributes to the theoretical studies are the future studies could be done in this area.

Chapter 1

Theoretical Background

1.1 Argument Structure

Language comprehension in people normally occurs in a quite short time and with very little if any delay. (Marslen-Wilson, 1975). This wisdom underlines that language processing is involved with deciphering a message and its participants in a very short time span. In linguistic terms, this structure is called argument structure.

The concept of argument structure is not one of those abstract notions that we do not deal with in our daily life. In simple words, in natural languages, when a structure is made, one element selects its required elements to complete its message. The following examples are from the Pustejovsky and Batiukova (2019), and a brief review of its relevant concepts are illustrated in the following.

Considering the following simple equations:

- (1) a. $\frac{1}{2} = 0.5$
- b. $8 > 5$

As is evident in (1), the message this equation is trying to convey is described by the mathematical functions of = and >. In addition, the numerical arguments are on the different sides of the equation because they fit into the logical statement of the functions. Therefore, it can be assumed that they have been chosen based on the needs of the message and the meaning of the functions.

Considering this simple algebraic structure, in linguistic utterances as in (2), it can be observed that in terms of structures, there are some similarities.

- (2) a. The book is interesting.
- b. Peter attended MIT.

In (2a), the adjective is the element that needs a nominal component to complete its function. In addition, in (2b), the verb, **attend** has two argument slots which are needed to be filled. However, in (2b), the structure is a bit more complicated because the arguments are chosen based on the event the verb is describing. For example, in (3), although the structure in both of the examples is the same, in terms of the semantic relationships between the elements, they are different. Evidently, the second argument in (3a) is the GOAL or the final part of this transportation; while, in (3b) it is the SOURCE or the starting point.

- (3) a. The train arrived in Venice.
 b. The flight departed from Madrid.

It can be assumed that the notion of argument selection is similar to an equation, where the different sides of it are chosen to convey a message based on the function or selecting element. In linguistic terms, these selecting elements can be verbs, nouns, adjectives, and prepositions (Pustejovsky and Batiukova, 2019).

The argument structure in verbs is highly related to the events they describe; in other words, the argument structure has to map the relationship between participants of an event, the way they interact, and the number of participants. Linguistically speaking, thematic or semantic roles provide the mapping among the participants, and the verb valence considers the number of participants required in the event.

This research intends to study the thematic roles of the Complex Predicates in Persian. In this sense, these complex verbs are the elements that describe the events. The event's nature and our understanding determine which participants are encoded as verbal arguments, how they interact with one another, and what roles they play in the event. By way of illustration, what thematic role they allocate.

In contrast with the simple constructions, these verbs are made of two elements. The role each of these elements plays in determining the thematic roles is controversial among scholars. Precise mapping of the roles and the effect of these elements can shed light on the way speakers of these languages comprehend these constructions and how they produce them. Namely, when they are making a complex predicate with an agent in the subject, how do they choose the elements of the complex predicate, and how their understanding of the roles changes by changing each of the elements.

In the following sections, the concept of complex predicates and different proposals on their argument structure has been illustrated; next, the notion of thematic roles is discussed; followed by the cognitive account of the thematic roles and a review of several related studies; eventually, the niches and the research questions

are discussed.

1.2 Persian Complex Predicates

Persian as a verb-final language, constructs its unmarked transitive sentences in an SOV order as in (4):

- (4) a. Reihane hadiye-ash-ra did
 Reihane her.gift-ra saw.3SG
 "Reihane saw her gift."
 b. Ali puzzle-ra hal kard
 Ali puzzle-ra solve did.3SG
 "Ali solved the puzzle."

Comparing the examples in (4), it can be seen that both of the constructions are transitive sentences, with a direct object which has been marked by -ra. In addition, the comparison indicates that in the first sentence, the verb is made of just one element, while in the latter, the verb is made of two components. In utterances such as (4b), which contains two components in their verbs, one of the elements is called a *Light verb*, which can be defined as a commonly used verb that has lost its main meaning and is incorporated for making constructions; and a *Non-Verbal* or [Goldberg et al. \(2003\)](#) representing a construction-based approach, calls them *host*. As it can be seen in (5), these elements can be chosen from different phrasal categories; namely, nouns, adjectives, particles, prepositional phrases, and complex nouns.

- (5) a. xar kardan/fodan
 donkey doing/becoming
 "to fool or become fooled"
 b. sabok kardan/fodan
 light doing/becoming
 "to degrade"
 c. birun kardan
 out doing
 "to dismiss"
 d. be yad daftan
 to memory have
 "to have in mind."
 e. dast o pa zadan
 head and foot hitting
 "to try"

In the literature these constructions are mostly either called a Complex Predicate construction (CPr) (Karimi, 1997, among others) or light verb construction (LVC) (Karimi-Doostan, 2005, among others). Notwithstanding the abundant terminology that has been proposed for these constructions, we will refer to such combinations as Complex Predicates (CPr), and for the first component of them, we use non-verbal (NV) since it has been more frequently used in the literature.

Complex predicates are one of the most productive constructions in Persian. The diachronic studies report that this trend began in Middle Persian (224-651 A.D.) and became dominant in Early New Persian, Dabir-Moghaddam (1997) reports from Rastorgueva that the number of verbs in that period was not so high; while most of the verbal constructions had the noun plus verb construction. In most cases, *kartan* 'do' constitutes the verbal part in the combination, e.g., *zen kartan* 'to saddle,' *rofnih kartan* 'to light,' and *azmayef kartan* 'to test.'

In addition, the analysis of the loan words in Persian indicates the same trend. Dabir-Moghaddam (1997) mentions from Sadeghi (1993) that although, a limited number of borrowed verbs from Arabic were used as simple verbs (e.g., *bal?idan* 'to swallow; to devour' from Arabic *bal?* *raqsidan* 'to dance' from Arabic *raqs*, *fahmidan* 'to understand' from Arabic *fahm*), the majority of borrowed verbs were used as complex. Such as all new borrowings from western languages (e.g., *faks kardan/zadan* 'to fax', *telefon kardan/zadan* 'to telephone', *montaz kardan* 'to montage', and *google kardan* 'to google'). In addition, many of the existing simple verbs have compound counterparts that are derived from the simple verb itself, which in most of the cases, the simple counterpart is considered as the more formal variation (e.g., *Jang kardan* from *jangidan* 'to fight', *gerye kardan* from *geristan* 'to cry', and *qalt zadan* from *qaltidan* 'to roll'). It is worth mentioning that, these days, due to the impact of technology on writing and the limitation of characters, a limited number of verbs have undergone the reverse of the trend. A simple form of them has emerged from their complex counterpart, such as *zangidan* from *zang zadan* 'to call.'

As the literature indicates, Complex Predicate constructions have attracted the attention of many scholars. One of the most significant distinctions between the studies considers the compositionality of the CPr elements. In the analyses of Karimi (1997); Folli et al. (2005), the meaning of these constructions is considered to be derived post-syntactically and are accounted as fully-compositional constructions. In Karimi-Doostan (1997), the same approach has been taken; however, in the latter, the construction is formed based on the meaning of the non-verbal, since the LV is considered to be semantically bleached; while in the former, the verb and the verbal arguments are inserted into the syntactic construction.

On the other hand, scholars such as Goldberg (1995) and Samvelian and Faghiri (2014) propose a posterior-compositional approach toward the CPrs. They argue that from an *a posteriori* analysis point of view, CPrs can be treated the same as the idiomatic constructions; in this regard, they argue that each of the CPrs corresponds to a construction of the pairing of meaning and form; namely, the CPrs' meaning is considered as a whole regardless of its components. Moreover, they argue that the CPr-constructions can be grouped into a more general group based on their semantic and syntactic features, and as a final result, they can be grouped into a more generalized network.

Considering the mentioned approaches and the studies in the literature, before moving towards the details of the studies, let us review the notion of the LVs and observe how the Light Verbs are accounted for and discussed in the literature or, more specifically, in from the perspective of Hale et al. (1993) and Grimshaw and Mester (1988), thereupon I will discuss the constructionalist analyses of Folli et al. (2005), and the lexicalist approach of Karimi-Doostan (2005), and the construction-based analysis of Samvelian and Faghiri (2014). I will illustrate how they account for the allocation of the thematic roles and what possible answers they provide to our research question.

1.2.1 Approaches towards Light Verbs

Grimshaw and Mester (1988)'s approach towards Light Verbs depict them as a type of verb which has lost its semantic load; therefore, they are not able to form an argument structure independently; and they need to combine with another element, which is the NV, to form a complete argument structure. Consider the famous Japanese example of *suru* in the examples from Grimshaw and Mester (1988) in (6).

- (6) a. John-ga Bill-to HANASHI-o shiteiru
 John-Nom Bill-with talk-Acc suru
 "John is talking to Bill."
 b. Ya-ga mato-ni MEICHUU-o shita.
 arrow-Nom target-to hit-Acc suru
 "The arrow hit the target."
 c. America-ga 200-nen-mae-ni Igrisu-kara DOKURITSU-o shita.
 America-Nom 200-year-ago-at England-from independence-Acc suru
 "America became independent of England 200 years ago."

As illustrated in the examples, although *suru* is present in all of the constructions, their argument structure is not fixed. As is evident in (6b) and (6a), the arguments

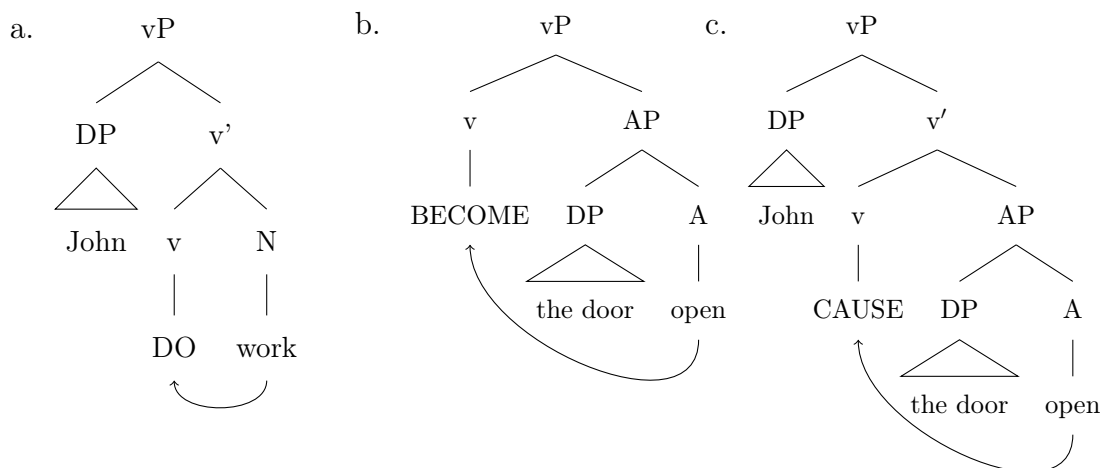
do not seem to have similar semantic features, e.g., in the former, the role of John is agentive, and the role of the arrow in the former is more or less patient-like, while the light verb *suru* is the verb in both of the examples.

Light verbs seem to be so bleached that they are not able to affect the number of arguments as well in the argument structure. This can be inferred from the comparison of (6c) with (6a) and (6b), which evidently have a different number of verbal arguments. The interpretation can be acquired by the comparison of the Persian light verbs in the CPrs. Karimi-Doostan (2005) states that the LVs are excessively used in languages such as Persian and Kurdish. Consider the following examples of the verb *kardan* from Karimi-Doostan (2005) in Persian. As it can be apparently observed in (7), although the LV is fixed in all of the utterances, the valency of the verb has changed from ditransitive as in (7a) to intransitive in (7c).

- (7) a. John aroosak-ra be Mary ehda: KARD
 John doll-ra to Mary giving did.3SG
 "John gave the doll to Mary."
 b. Columbus amrika-ra kaff KARD.
 Columbus america-ra discover did.3SG
 "Columbus discovered America."
 c. John narmeʃ KARD.
 John exercise did.3SG
 "John exercised."

Another approach towards LVs, which is mostly followed by a constructionalist approach towards the CPrs such as Folli et al. (2005); Megerdooimian (2001), considers LVs as more abstract notions. Hale et al. (1993) in a radical approach towards argument structure argue that verbs, even in English, are not syntactically simplex, while they are made of a light verb and a non-verbal element. Therefore, in the final structure, the verb is the result of the incorporation of these elements. For instance, in languages such as English, the light verb is a zero morpheme in most cases, and only the non-verbal component could appear in the final structure. However, in cases where verbalizing occurs through the addition of suffixes such as *-en* (*enlighten*), *-ize* (*criminalize*) or *-ify* (*clarify*), this zero morpheme of LV becomes overt. Following this hypothesis, in (8a), the verb has the underlying unarogative meaning of "do work." Moreover, as (8b) and (8c) indicate the same root in different syntactic structures, it consequently takes two different LV interpretations, namely inchoative and causative.

(8)



These approaches towards the light verbs have established the foundations for different accounts toward the complex predicates, which I will review in continue section 1.3. As mentioned earlier, Complex Predicates have been extremely used in Kurdish, Persian, and Korean languages. Although it would be difficult to have a complete list of the LVs in such languages, it can be argued that almost 20 of them are being used today in Persian. In table 1.1, from one of the most recent researches on the classification of LVs [Karimi-dustan and Eshaghi \(2019\)](#) shows a relatively complete list of the LVs of Persian, based on their heavy meaning in [Levin \(1993\)](#) verb classification.

Among the mentioned LVs in table 1.1, the light verb *kardan* 'to do' has almost entirely lost its heavy interpretation and is the most productive LV in Persian. In addition, the LV *shodan* 'to become' is systematically used in so-called passive or unaccusative constructions.

1.3 Approaches towards the Persian Complex Predicates

As was mentioned in the previous sections, the analyses of the CPRs can be categorized into two main classes firstly, scholars who consider them as compositional and have studied the elements of CPRs as syntactically independent and *a priori*; and scholars who investigate them virtually like the idiomatic constructions and have proposed an *a posteriori* analysis of them. The following sections have reviewed different approaches toward the CPRs based on the lexicalist, constructionalist, and usage-based approaches.

Verb Class	Light Verb
State Verbs	daftan 'to have'
Do Verbs	kardan 'to do'
Verbs of Creation and Transformation	ʃhodan 'to become' gaftan 'to turn'
Verbs of Putting	gozaftan 'to put'
Verbs of Existing Force	kefidan 'to pull'
Verbs of Preception	didan 'to see'
Verbs of contact by impact	zadan 'to hit'
Verbs of change of possession	dadan 'to give' baxfidan 'to give'
Verbs of Obtaining	gereftan 'to get' jaftan 'to obtain'
Tape Verbs	bastan 'to tie'
Motion Verbs	?a:madan 'to come' ?a:vardan 'to bring' residan 'to arrive' raftan 'to go' ?oftadan 'to fall'
Verbs of Throwing	?anda:xtan 'to throw'
Verbs of Sending	bordan 'to take'
Verbs of Contact	xordan 'to collide'

Table 1.1: Semantic categorization of Persian Light Verbs proposed by [Karimi-dustan and Eshaghi \(2019\)](#)

1.3.1 Lexical approach

Perhaps one of the first attempts in systematically studying the contribution of the CPRs counterparts, e.g., light verb and non-verbal elements, is orchestrated by [Karimi-Doostan \(1997\)](#), who follows [Grimshaw and Mester \(1988\)](#)'s perspective on LVs. In his research, Karimi-Doostan takes a lexicalist approach toward the CPRs and defines the LVs as notions that cannot assign theta-roles and form an argument structure. Accordingly, the LV is required to combine with a non-verbal element for developing a complete verb in terms of semantic and syntactic features. From this perspective, they claim that meaning plays an integral role in determining the number and the types of arguments. Considering the characteristics of the LVs in the standpoint of [Grimshaw and Mester \(1988\)](#), Karimi-doostan highlights that

both of the critical attributes of the argument structures are denoted by the NVs.

In this approach, the LV is supposed to assign aspectual properties, while the argument structure formation is done by the non-verbal. He supports his argument by analyzing the role of the NVs and the LVs in forming the verb valency of the CPrs. An example of this was presented in (7), where although the light verb is fixed, the valency of the verb ranges from three arguments to one, governed by the meaning of the NV element.

They continue their debate on the thematic roles in the CPrs, indicating that although in (9) the LV has changed, the thematic role of the subject has remained unaffected. Where although the verb has changed, the thematic role is remained fixed.

- (9) a. John fot kard
 John death did.3SG
 "John died"
 b. John fot fod
 John death become.3SG
 "John died"

In order to test the role of these constructions, let us add *amdan*, 'intentionally' to them, which indicates if the role of the arguments is an agent or not. In order to draw the comparison, consider the examples in (9), and (7c).

- (10) a. #John amdan fot kard/fod
 John intentionally death did.3SG/became.3SG
 "#John died intentionally"
 b. John amdan narmejkard
 John exercise did.3SG
 "John exercised intentionally"

As is evident in (10), the first structure does not make sense regardless of its LVs, while in the second example (10b), the role changes because of the meaning of the construction, denoted by the NV. Based on the observation in the above examples, Karimi-Doostan (2005), states that, a critical point in the perspective of Hale et al. (1993); Hale and Keyser (2002), illustrated in (8a) and (8c), and (8b) is that the *v* dominates the *vP* in constructionist approach, and its existence is related to the presence or absence of an external argument, which by its presence it implies having *v*, which is a sign of accusative case. In this regard with the examples in (10) and (7), they argue that the LVs do not account the same of the explicit *v*, argued by Folli et al. (2005).

Considering the semantically bleached notion of the LVs, Karimi-Doostan performs his analysis according to the NVs. He continues that the CPrs constructions can be categorized based on the NV combined with, namely, compositional and non-compositional CPrs. He defines a compositional CPr when the NV is a predicative transparent noun and displays an argument structure; he argues that in compositional constructions, the meaning of the whole construction is closer to the meaning of each of the elements of the CPr as in (12). On the other hand, he explains that if the noun is thematically opaque and does not constitute an argument structure, a non-compositional CPr is made, and in contrast with compositional type, their meaning is more idiomatic and cannot be inferred based on the CPrs counterparts; examples of this can be seen in (11).

(11) a. *guʃ kardan*
 ear to-do
 "to listen"

b. *sar zadan*
 head to-hit
 "to visit"

(12) a. *motaleʔe kardan*
 study to-do
 "to study"

b. *?anjam dadan*
 performing to-give
 "to perform"

In Karimi-Doostan's perspective, the LVs have the role of encoding the present/-past, case and agreement features, and the aspectual properties of the constructions. He proposes three categories according to the aspectual properties of the LVs, namely, initiatory, transition, and stative aspectual roles.

Initiatory verbs such as *dadan*, 'to give' form CPrs with at least one external argument, that is, either unergative or transitive (or causative) CPrs. They are thus compatible with nouns having at least one external argument, which refers to the initiator of the action denoted by the CPr and is associated with the aspectual role Initiatory assigned by the verb. Transition verbs, such as *xordan*, 'to hit' form CPrs with a single internal argument, or, in other words, unaccusative predicates. They are thus compatible with nouns having at least one (internal) argument. The latter is mapped into the subject function and receives the Patient role. Consider the NVs, *ʃekast*, 'defeat' and *kefik*, 'guard, watch' in the following examples.

- (13) a. Ali Sasan-ra jekast dad
 Ali Sasan-ra defeat gave.3SG
 "Ali defeated Sasan"
- b. Ali hame-ye jab-o kejik dad
 Ali all-Ez night-ra watch gave.3SG
 "Ali was on watch throughout the night."

As the mapping in (13) indicates, in (13a), the verb allows two thematic roles, e.g., a patient and an agent; while in (13b), the verb allows only for one role. As a result, (13a) can also be compatible with a Transition verb as in (14), while the latter only allows the initiatory verbs.

- (14) Sasan jekast xord.
 Sasan defeat collide.
 "Sasan was defeated."

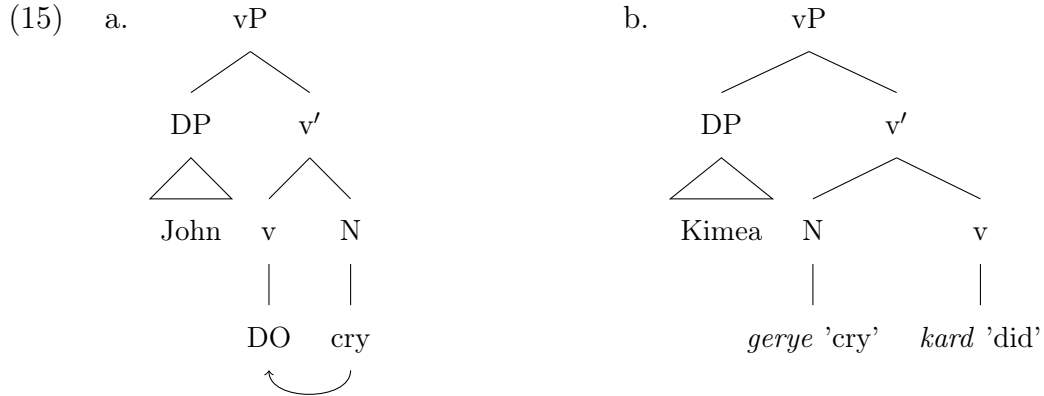
The lexicalist approach towards the CPrs highlights the spots of contrast among the characteristics of the counterparts of the CPrs. The intuition that can be derived from this approach is that, in the examined cases, the meaning of the NV governs the event and argument structure of the CPrs. The lexicalist approach proposes a potential answer to our research question. In this frame, the thematic role of the CPr constructions is highly related to the meaning and the notion of the event, determined by the non-verbal elements.

1.3.2 Constructionist approach

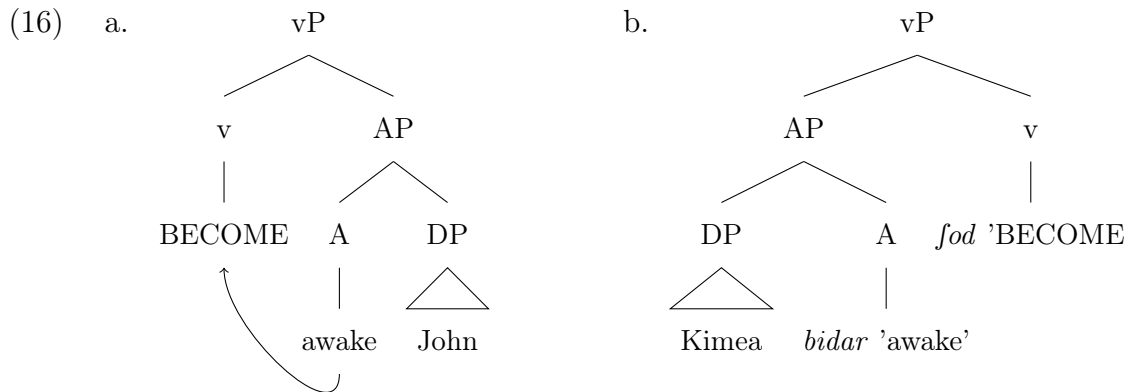
Another approach towards the CPrs is rooted in the framework of [Hale et al. \(1993\)](#); [Hale and Keyser \(2002\)](#) on light verbs. As was mentioned earlier, this framework explains that the basis of the semantic properties in construction is the LV, and the other elements are positioned into the syntactic structure. Considering [Hale et al. \(1993\)](#); [Hale and Keyser \(2002\)](#), unergatives are formed when a nominal element is incorporated into a light verb which selects for an external argument. Similarly, inchoatives are formed when an adjectival element is incorporated into a light verb that does not select for an external argument. Observing this approach, not only in languages that have these elements explicitly, such as Persian, but also in languages such as English that do not have the LVs on their surface, in most of the cases, the LV is present and forms the event.

Scholars such as [Megerdooian \(2001, 2012\)](#) and [Folli et al. \(2005\)](#) follow this framework for analysing the CPrs. In this approach, in contrast with those of [Karimi-Doostan \(1997\)](#), the emphasis has altered from the lexicon, namely the NV

element, to the syntax, and it is the LV that designates which roles and elements are needed for the formation of the event. (15a), the following examples indicate the verb *gerye kardan*, 'to cry' in English and Persian, from Hale and Keyser's perspective:



Following [Folli et al. \(2005\)](#)'s approach, and as the comparison of the examples in (15a) and (15b) indicate, the LVs are in the position of the head of the *v*. Therefore, they argue that in cases where the CP allows for the external structure is because of the argument structure and the LVs; while in cases where the CP does not receive the features of an Agent role, it is because the LV is not allowing the external argument selection as in (16):



Following this account of the CPs, [Folli et al. \(2005\)](#), propose a number of features related to the NV element and the LVs. In this respect, they argue that the Agentivity and Causativeness, State and events, and Duration are intrigued by the LVs, while the NV account for the Aktionsart features of the constructions.

In this framework, the Agentivity and Causativeness appear to be motivated by the LVs in the sentences. Considering the examples in (17), it is evident that the shift from an agentive to a non-agentive structure is motivated by a change

in the light verb. In other words, the sentence has shifted from a causative to an inchoative argument structure due to the agentive features of *dâdan* ('give') in (17a) to inchoative because of the features of *xordan*, 'collide' in (17b).

- (17) a. tim-e mâ unâ-ro fekast dâd
 team-EZ we they-râ defeat gave.3SG
 "Our team defeated them."
 b. tim-e mâ az unâ fekast xord
 team-EZ we of they defeat collided.3SG
 "Our team was defeated by them."

They support their claim by adding an agentive adverbial such as *amdan* 'intentionally', which indicates that in cases where the subject is agentive, its role will remain constant even when the NV element's category is manipulated, while the patientive structure turns to be ungrammatical as in (18).

- (18) a. Kimea amdan dâd zad
 Kimea intentionally yell hit.3SG
 "Kimea yelled intentionally."
 b. *ghazâ amdan dast xord
 food intentionally hand collided.3SG
 "Food became intentionally touched."

They continue their argument that the only case where the LV does not project the agentive or patientive features is related to the verbs of motion such as *gozasht*, 'to pass' and *amad*, 'to come' which in their light verb meaning are affected by the NV element, while in their heavy meaning they have an agentive meaning. Considering the examples in (19), it can be seen that although with the presence of the adverbial *amdan*, 'intentionally' the examples with the heavy form of these verbs is agentive (19a), their light meaning counterparts will turn to be ungrammatical (19b).

- (19) a. Kimea *amdan* az xiyâbun goza[ht
 Kimea intentionally of street passed.3SG
 "Kimea intentionally crossed the street."
 b. *Kimea *amdan* dar gozasht
 intentionally away passed.3SG
 "Kimea intentionally passed away."

They argue that this exception in the behavior of the LVs is perhaps rooted in the characteristics of the motion verbs since they change between an agentive/unergative and an inchoative/unaccusative reading.

The other feature they consider to be bonded to the LVs of the CPrs is stativeness and eventiveness. They discuss that some of the LVs, such as *datan*, 'to have' are only grammatical in a stative form, and they are unable to form progressive constructions. They support their claim by presenting an example of the contrast between *daftan*, 'to have' and *avardan*, 'to bring' in (20), that by having a fixed NV element, *jad*, 'memory', the construction with *avardan* can be either eventive or stative and grammatical; while for *daftan* the only the stative form is grammatical.

- (20) a. Kimea esm-e un-o be yâd dâr-e
 Kimea name-Ez her-râ to memory have.3SG
 "Kimea has her name in her memory."
 b. *Kimea esm-e un-o dâr-e be yâd dâr-e
 Kimea name-Ez her-râ have-3sg to memory have.3SG
 "*Kimea is having her name in her memory".
 c. Kimea esm-e un-o be yâd mi-yar-e
 Kimea name-Ez her-râ to memory hab-bring.3SG
 "Kimea remembers her name."
 d. Kimea esm-e un-o dâr-e be yâd mi-yâr-e
 Kimea name-Ez her-râ have.3SG to memory hab-bring.3SG
 "Kimea is remembering her name."

Another characteristic of the CPrs, which [Folli et al. \(2005\)](#) and [Megerdooimian \(2001\)](#) claim to be associated with the LVs, is the change in the duration of the events. They state that in the same event, different LVs are at the helm of the duration of the CPrs, namely, a verb such as 'to touch' can be formulated by two LVs, *zadan*, 'to hit', which implies punctuality, and *kefidan*, 'to pull' which implies duration.

In order to understand the role of the NVs in the CPrs from a constructionist approach, it is crucial to have a better look at the correlation of the embedded arguments of the *vP* and the Aktionsart of the whole predicate. Considering the examples in (21), it can be seen that in terms of Aktionsart, (21a) is denoting an *Activity*; while, in the other examples the predicates are characteristically *Accomplishments*.

- (21) a. John workd.
 b. The door opened.
 c. John opened the door

[Hale et al. \(1993\)](#); [Hale and Keyser \(2002\)](#)'s construtionalist approach towards the event structure, the crucial difference between the two classes seems to be the

type of clause that appears in the complement of *v*. In other words, considering the trees of the above examples, in the situations where the verb denotes a *telic* Accomplishment such as (8c), and (8b), the lower phrase is a predicate, and its subject indicates a change of state. While in the activities such as (8a), the whole predicate indicates an agent, and the lower phrase incorporated into the verbal shell is a nominal expression.

Folli et al. (2005) argue that the NVs are to be accountable for the Aktionsart characteristics of the whole predicate in Persian CPrs. They support their claim by an example of the verb *awake* in transitive, *bidar kardan* and intransitive *bidar sodan*, modes, and to cry *gerje kardan*. They test the aktionsart of the examples by adding the *for an hour* displaying an atelic activity, and (*in*) *an hour* revealing a telic accomplishment. As the examples indicate, the telicity was not affected by the change of the LVs in (22a) and (22b); while by having *kard*, to do, fixed and changing the NV to *gerje*, to cry, a change in the telicity can be detected.

- (22) a. Kimea ye sâ'ate/* barâye ye sâ'at bidâr shod
 Kimea one hour/for one hour awake became.3SG
 "Kimea became awake within an hour."
- b. Kimea ye sâ'ate/ *barâye ye sâ'at Papar-ro bidâr kard
 Kimea one hour/for one hour Parpar-râ awake did.3SG
 "*Kimea is having her name in her memory".
- c. Kimea *ye sâ'ate/ barâye ye sâ'at gerye kard
 Kimea one hour/for one hour cry did.3SG
 "Kimea cried for one hour."

By comparing and contrasting different NV types, they expand their idea on the role of the NVs and argue that all of the NV types denote telic events; however, NVs, which are non-eventive nouns such as *dad* in *dad zadan*, 'to yell', denote atelic features. A summary of their typology and examples of them is reported in table 1.2.

Telic Complex Predicates	
Prepositional Phrase + LV	<i>be donyâ âmadan</i> (to world coming) 'to be born'
Particle + LV	<i>dar gozâftan</i> (away passing) 'to pass away'
Adjective + LV	<i>derâz kefidan</i> (long pulling) 'to take a nap'
Eventive Nominal + LV	<i>fekast xordan</i> (defeat colliding) 'to be defeated'
Atelic Complex Predicates	
Nominal + LV	<i>dâd zadan</i> (scream hitting) 'to yell'

Table 1.2: Telicity of the Persian Complex Predicates according to Folli et al. (2005)

To sum up, the constructionalist approaches toward the Persian Complex predicates follow a compositional analysis. Namely, they investigate the constituting elements of these constructions independently and, based on their behavior, describe the characteristics of CPrs. In the studies of [Folli et al. \(2005\)](#), and [Megerdooomian \(2001\)](#), by following the theory of [Hale et al. \(1993\)](#); [Hale and Keyser \(2002\)](#), it was argued that the constituting elements are inserted into the verbal structures; in this regard, they claimed that the Aktionsart characteristics of the CPrs are interpreted according to the NV elements. In addition, it was claimed that the Agentivity/Causativity, eventiness, and the duration of the events are motivated based on the attributes of the LV.

Considering the topic of the current study, one hypothesis that can be derived from the constructionalist approach towards the CPrs is that the LVs, as the head of the *vP*, are able to affect the thematic roles of the constructions, and different LVs, such as *zad*, to hit, and *xord*, to collide, are able to denote different roles and levels of agentivity.

1.3.3 Construction-based approach

In contrast with the lexicalist and constructionalist approaches, which view the CPrs as radically compositional, the construction/usage-based theory states that a fully-compositional approach fails in explaining several important issues about the characteristics of the CPrs. They suggest that to tackle these drawbacks, it is necessary to establish the interpretations based on the context of the LVs and the NVs.

Scholars from the usage/construction-based background who study the CPrs, based their argument on the necessity of observing these roles from a *a posteriori* point of view. They begin their analysis by emphasizing the deficiencies of the compositional approaches. As [Samvelian and Faghiri \(2014\)](#) explain, the fully compositional approaches towards the CPrs are not able to account for several critical issues. Namely, they argue that the CPr elements do not demonstrate a consistent contribution in all of the instances, their semantic content is not predictable, and a collocational association between the components cannot be accounted for compositionally.

In the following, the examples regarding the deficiencies have been presented and argued, and in the last part of this section, the principles of the posterior analysis have been discussed.

[Samvelian and Faghiri \(2013a\)](#) following [Pollet \(2012\)](#), begin their criticism with the problem of the non-consistent contribution of the CPr counterparts. They

argue that, in terms of the LVs, the characteristics that Folli et al. (2005) assign to the LVs. e.g., Agentivity/Causativity, eventiveness, and the duration of the events can be challenged considerably. One of the LVs that in Folli et al. (2005) was analyzed and emphasized on was *zadan*, 'to hit', which was argued to select agent in the CPrs and demonstrate a punctual duration of action. However as it can be seen in (23), Samvelian and Faghiri (2014), asserts that this LV is able to form constructions which are not agentive, and by adding the adverbial *amdan* 'intentionally', which was the test provided by Folli et al. (2005), they become non-sense or ungrammatical.

- (23) a. in nān kapak zad-e ast
this bread mold hit.3SG be.PRS.3SG
"This bread has gone moldy."
- b. #in nān amdan kapak zad-e ast
this bread intentionally mold hit.3SG be.PRS.3SG
"#This bread has gone moldy intentionally."
- c. āb jax zad
water ice hit.3SG
"The water froze."
- d. #āb amdan jax zad
water intentionally ice hit.3SG
"#The water froze intentionally."

In addition, in terms of the NV elements, Samvelian and Faghiri (2014), argue that, in almost all of the phrasal categories of the NVs which Folli et al. (2005) classified (can be seen in table 1.2), there are some inconsistencies in terms of their Aktionsart. As an example Folli et al. (2005) argued that the Adjective NVs are always telic, while as it can be seen in the CPr made of the adjective, *penhan*, 'hidden', incorporated with *daftan*, 'to have' in fact when the construction co-occurs in an atelic context it makes more sense (24b).

- (24) a. maryam xabar=rā az omid penhān dāft
Maryam news=ra from Omid hidden have.3SG
"Maryam hid the news from Omid."
- b. maryam xabar=rā se sāl az omid penhān dāft
Maryam news=ra three year from Omid hidden have.3SG
"Maryam hid the news from Omid for three years."
- c. #maryam xabar=rā dar yek sâ?at(yek-sâ?ate) az omid penhān
Maryam news=ra in one hour from Omid hidden
dāft
have.3SG

(Lit)"#Maryam hid the news from Omid in one hour."

In other examples, it has been demonstrated that the NVs from the prepositional phrase category, which were also classified as telic, make sense in atelic contexts as well in (25). Moreover in the non-eventive-noun category, the same deficiencies can be observed (26) (Samvelian and Faghiri, 2013b).

- (25) a. maryam omid=rā be masxare gereft
Maryam Omid=ra to mocking take.3SG
"Maryam mocked Omid."
- b. maryam omid=rā (bara-je) sāʔat-hā be masxare gereft
Maryam Omid=ra (for) hour-PL to mocking take.3SG
"Maryam mocked Omid for hours."
- c. #maryam omid=rā dar yek sāʔat(yek-sāʔate) be masxare gereft
Maryam Omid=ra in one hour to mocking take.3SG
(Lit)"#Maryam mocked Omid in one hour."
- (26) a. mār pust andāxt
snake skin throw.3SG
"The snake sloughed off."
- b. #mār (baraje) do ruz pust andāxt
snake (for) two day skin throw.3SG
(Lit.) "#The snake sloughed off for two days."
- c. mār dar do ruz pust andāxt
snake in two day skin threw.3SG
"The snake sloughed off in two days."

In addition to the problems of the inconsistency of the characteristics of the CPr counterparts, as it was also accounted for in the lexicalist approach, all of the CPrs do not seem to be interpretable from their elements. In the lexicalist perspective, Karimi-Doostan (1997, 2005), argued for the opaque and the transparent CPrs; in the latter, the meaning of the whole unit can be understood based on the CPr elements, while in the former, the same as the idiomatic expressions the meaning cannot be comprehended compositionally (see section 1.3.1). In addition, as Goldberg (1996) explains, viewing the components of a CPr such as *dust daftan*, literally 'friend have', one might suggest that it means *to have a friend*; however, in a different manner from the meaning of its elements it means 'to love/to like'.

The last challenge posed to the compositional approaches of the CPrs considers the collocational association between the CPr components. Namely, in cases when the synonymous NVs of the CPrs are combined with different LVs to construct a synonymous CPr. For instance Samvelian and Faghiri (2013a) argues that, although

both *hesadat* and *rafk* in Persian have the meaning of 'envy', the incorporate with different LV, successively, *bordan*, 'to take' and *kardan*, 'to do'. From a compositional point of view, since the change in the LV is always rooted in syntactic reasons, it is almost impossible to account for such combinations.

Building upon the challenges proposed to the fully-compositional approaches towards the CPrs, scholars who follow construction-based frameworks such as Goldberg et al. (2003) and Samvelian and Faghiri (2013a, 2014), argue for a *postapriori* analysis of the CPrs. They based their argument upon the idea that the fully-compositional understandings of the CPrs are not able to propose a comprehensive analysis of them. In a construction-based approach, it has been argued that CPrs need to be treated as idiomatic expressions. As Samvelian and Faghiri (2013b) explains for the ideas of Nunberg et al. (1994), on idioms, the elements of the CPrs assigned a meaning based on the context in which they stand in. Therefore, CPrs are compositional in the sense that their meaning is distributed among their counterparts; and yet idiomatic because the contribution of each member cannot be analyzed in isolation. In this regard, scholars of Construction-based approaches argue that one of the sophisticated privileges of this account of the CPrs is that it enables the researchers to provide more fine-grained classifications of the components of the CPrs and investigate their productivity.

As noted in the introduction of the section, in the construction-based approach, there are three principles for the analysis of the CPrs. Firstly, they need to be considered part of a Construction; secondly, they need to be grouped into different classes based on their semantic and syntactic characteristics, each class representing a partially fixed construction. Furthermore, they are mapped in networks representing their different semantic and syntactic relations between CPrs.

In this approach, CPrs are treated with respect to their independence and compositionality while they are needed to be considered as a part of a construction. The definition of the construction in this framework is different from what we have dealt with until this section. As Goldberg (1995) explains, a construction is a conventional association between form and meaning, (*form - meaning pairings*). In other words, the CPrs are considered to be stored with their conventional meaning and, in this sense, are comparable to the lexemes. A CPrs, in this regard, can be assigned roles and linguistic features, but with respect to the context, they appear in. Samvelian and Faghiri (2013b) reporting from Pollet (2012), proposes a sample of these constructions for *zadan*, 'to hit', classifying it in Spreading-*zadan*, Instrument-*zadan*, Slandering-*zadan*, and forming-*zadan* constructions. The example of this construction representation is demonstrated in the table 1.3

The first line of the constructions indicates the general label assigned to the LV

Spreading-zadan construction	Ex: <i>rang zadan</i> , 'to paint'		
N0 Agent 'N0 applies N2 on N1' or 'N0 covers the surface of N1 with N2'	(be)N1 Ground	N2 Figure	V
Instrument-zadan construction	Ex: <i>mesvāk zadan</i> , 'to brush one's teeth'		
N0 Agent 'N0 accomplishes the typical action for which N2 is used (on N1)'	(be)N1 Patient	N2 Instrument	V
Slandering-zadan construction	<i>tohmat zadan</i> , 'to slander'		
N0 Slanderer 'N0 accuses N1 of N2'	be N1 Slanderee	N2 Slander	V
Putting-zadan construction	Ex: <i>tambr zadan</i> , 'to stamp'		
N0 Agent 'N0 puts N2 on N1' or 'N0 covers the surface of N1 with N2'	(be)N1 Ground	N2 Figure	V
Forming-zadan construction	Ex: <i>kapak zadan</i> , 'to mouldy'		
N0 Location 'N1 is formed on N0'	N1 Theme	V	

Table 1.3: Constructions of *zadan*, 'to hit' proposed by [Samvelian and Faghiri \(2013b\)](#)

constructions; the second line indicates the verbal valency of argument structure; namely, N0 represents the Subject, N1 stands for the Object, and so on; and the third line indicates the semantic roles and the last line shows the abstract meaning allocated to the construction.

Moving a step forward in the construction-based approach, the relationship of the constructions should be taken into account. A network of the connections of the constructions can be devised based on different semantic, such as hyponymy and hypernymy, and syntactic, such as valency alternations, relationships, they have.

In order to better illustrate the networks of the constructions let us review the networks of the constructions of *zadan*, 'to hit' proposed by [Samvelian and Faghiri \(2013b\)](#).

Considering the constructions of the *zadan*, 'to hit', it can be argued that some of the constructions are able to form a super-class and be grouped. For instance, as it can be seen in fig. 1.1, all of the constructions in which the N0 performs an action on N1, which is a considered as a surface, or in general on ground, such as the instrument-*zadan* and the putting-*zadan* construction in table 1.3, can be grouped and form a super-class (or become a sub-class) of *Locatum-zadan*.

The same approach can be used in terms of the semantic relationships among the constructions or their syntactic features. An example of these groupings and networks can be observed in fig. 1.2. Minding the hierarchical orders of the network from the top of it, at the second level, the type of the construction grouping is mentioned, e.g., different LVs, their verbal valencies, and the system class, which is an umbrella term for the super-classes. Next, different levels and types of them are

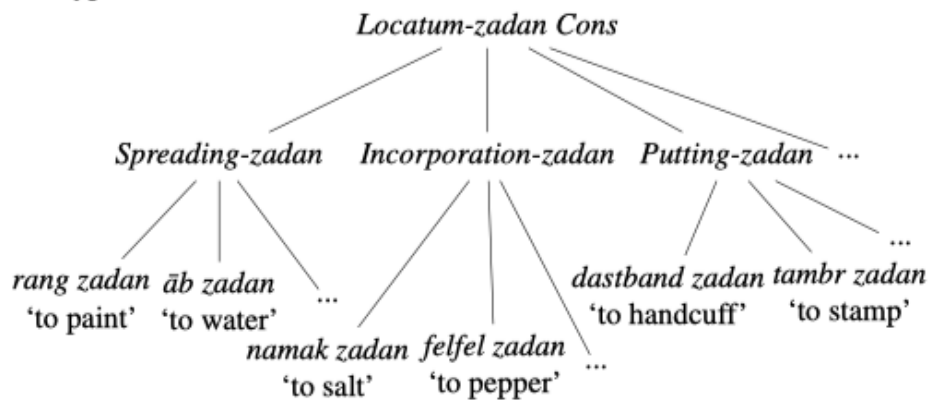


Figure 1.1: Subtypes of Locatum-zadan Construction

presented; for the matter of illustration, it can be seen that the network presented in 1.1 is located at the left side of the network, indicating the superclasses and sub-classes of locatum, instrument, and so on. Moreover, the active or passive mode of the CPRs is located under the valency term, and the LVs as a group are presented on the right side of the network. At the bottom of the network, it can be observed that the different CPRs are presented, being connected to their corresponding features in the network.

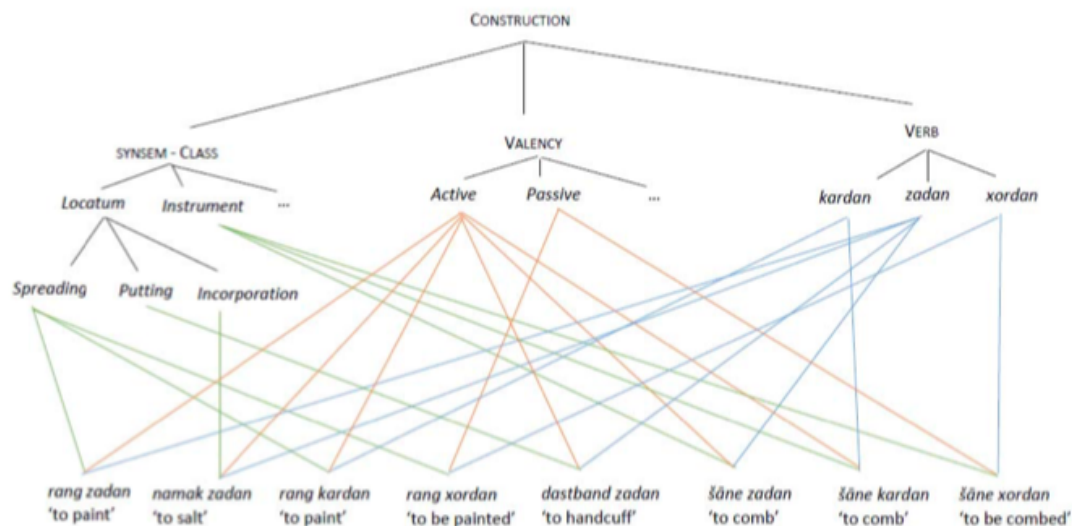


Figure 1.2: A partial hierarchy of Persian CPRs constructions proposed by Samvelian and Faghiri (2013b)

All in all, the review of the construction-based approach highlights some of the drawbacks of the aforementioned compositional approaches. In addition, the non-compositional approach towards the CPRs presents the question of how the

interaction of the LVs and NVs can vary among LVs and in different contexts or constructions. A more specific question regarding this research is how the allocation of the thematic roles can be formulated in this approach towards the CPrs, and whether there is a systematic pattern for each light verb or they are only randomly assigned to the NVs and based on their interaction the roles are allocated. Considering the impact of this approach on the current research, the non-compositional perspective proposes a third possible answer to the effect of the components of the CPrs in allocating the thematic roles of the CPrs. From this perspective, the answer is considered the interaction of both of the counterparts. In this regard, the thematic role will be the resultant of the interplay of CPr components.

1.4 Thematic Roles

Perhaps, thematic roles are one of the oldest and most studied conceptions of linguistics. One of the first readings of these notions was introduced in Panini's Sanskrit grammar almost twenty centuries ago. Even though it has been discussed a lot among scholars, there has been hardly a consensus on the interpretations of the thematic roles. This lack of agreement among the scholars can be rooted in two main reasons. Firstly, this notion has been studied and discussed in different linguistic frameworks, and this can lead to some fundamental differences in the interpretation (an example of these differences in frameworks can be seen in 1.4.2). The other reason that prevents the scholars from agreeing is the contrast between the traditional point of view, which prefers to offer lists of roles, and more recent approaches that prefer to offer clusters of roles. The first group, which, [Levin and Rappaport Hovav \(2005\)](#) call Semantic Role Lists, attempts to propose a discrete and atomic list of thematic roles. This line of research can be seen in the works of [Gruber \(1965\)](#), [Fillmore \(1968\)](#), and [Jackendoff \(1972, 1987\)](#). On the other hand, other scholars, such as [Dowty \(1991\)](#) and [Van Valin and LaPolla \(1997\)](#) have proposed a generalized view of these roles, namely Generalized Semantic Roles, by mainly avoiding lists and concentrating on describing the lexical features of the roles, consequently, proposing cluster roles.

Notwithstanding the differences, most scholars agree upon a definition based on the event structure and verbs argument selection. In this sense, verbs describe events, and based on the nature of that event, the participants of the verbs are encoded as verbal arguments. The fact that how these participants interact with each other and how they play their roles in the event defines the thematic roles

(Pustejovsky and Batiukova, 2019).

1.4.1 Case Roles

In Case Role Grammar, Fillmore (1968), considers thematic roles as notions highly related to the case system of languages. He developed his Case Role Grammar based on a debate on the position of the cases in the Generative Grammar. Some scholars of generative grammar believed that case is not a concept defined in the "deep structure" of languages. While it was seen as an inflectional "realization" of particular syntactic relationships which appear only in the "surface structure" (Chomsky, 1965).

However, Fillmore (1968) argues that the concept of the case, rather than being merely a morphological concept, is a syntactic notion. In his perspective, the case system in languages is a universal concept, defined in the deep structure. He argues that the basic structure of the sentences is made of a 'proposition' which is a tenseless set of relationships involving verbs and nouns. This proposition is later combined with the modality constituent to adjust the modalities of the sentence, e.g., negation, tense, mood, and aspect (27).

$$(27) \text{ Sentence (S)} \rightarrow \text{Modality (M)} + \text{Proposition (P)}$$

The Proposition constituent of the sentence can be expanded as in (28), where a verb and one or more case categories are selected based on their relationships with the verb.

$$(28) \text{ P} \rightarrow \text{V} + C_1 + \dots C_n$$

Case categories are made of a set of universal concepts, which are defined according to the understanding of the language speakers of the relations of the verbs and arguments. The following definitions of case roles in (29), are those derived from Fillmore (1968):

- (29)
- a. **Agentive**, the case of the typically animate perceived instigator of the action identified by the verb.
 - b. **Instrumental**, the case of the inanimate force or object causally involved in the action or state identified by the verb.
 - c. **Dative**, the case of the animate being affected by the state or action identified by the verb.
 - d. **Factitive**, the case of the object or being resulting from the action or state identified by the verb or understood as a part of the meaning of the verb.

- e. **Locative**, the case which identifies the location or spatial orientation of the state or action identified by the verb.
- f. **Objective**, the semantically most neutral case, the case of anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpretation of the verb itself; conceivably, the concept should be limited to things which are affected by the action or state identified by the verb. The term is not to be confused with the notion of the direct object, nor with the name of the surface case synonymous with the accusative.

It is worth noting that in Fillmore's perspective, these are some of the roles, but they are not all of them and their number is not fixed. Therefore, it is possible to add more relations to them if needed or to adjust them.

Fillmore's Case Role Grammar prepared the foundations for FrameNet Project (Baker et al., 1998), which is a lexical database based on Frame Semantics. In Frame Semantics, the meaning of specific lexical items is defined relative to a background frame or scene. In this regard, case roles were used as a part of these frames and labels indicating the relationship of the members of the predicate based on the meaning of the verb and regardless of their syntactic realization.

1.4.2 Lexical Conceptual Structure and Thematic Relation

Jackendoff introduces an organization for language, which comprises three autonomous levels of structure, namely, phonological, syntactic, and semantic/conceptual. These structures are made of several primitives and principles, which allow them to combine and form more complex expressions from the basic structures. In addition, this grammar also has a set of correspondence rules, which allows different levels to link between each other (Jackendoff, 1983).

Perhaps one of the most specific points in Jackendoff's perspective on the thematic roles is based on their role of them in the proposed language organization. As he argues in Jackendoff (1987), in contrast with the perspective of Fillmore (1968) and many other scholars, the thematic roles are not one of those correspondence rules between syntactic level and semantic/conceptual level or an interface notion; while they are structural configurations in the conceptual structure. Therefore, they can be defined by particular configurations of primitive operators. (30) represents the mapping of the conceptual structure level. As it can be seen, there are four main categories in this structure, namely, PLACE, PATH, EVENT, and STATE.

$$(30) \quad \text{a. PLACE} \rightarrow \left[\textit{PlacePLACE} - \textit{FUNCTION}(\textit{THING}) \right]$$

$$\begin{aligned}
\text{b. PATH} &\rightarrow \left[\text{Path} \left\{ \begin{array}{l} TO \\ FROM \\ TOWARDS \\ AWAY - FROM \\ VIA \end{array} \right\} \left(\left\{ \begin{array}{l} THING \\ PLACE \end{array} \right\} \right) \right] \\
\text{c. EVENT} &\rightarrow \left\{ \begin{array}{l} \left[\text{Event} GO (THING, PATH) \right] \\ \left[\text{Event} STAY (THING, PLACE) \right] \end{array} \right\} \\
\text{d. STATE} &\rightarrow \left\{ \begin{array}{l} \left[\text{State} BE (THING, PLACE) \right] \\ \left[\text{Event} ORIENT (THING, PATH) \right] \end{array} \right\}
\end{aligned}$$

Considering the expression "*under the table*", it can be said that, based on (30a), the element representing the basic category of PLACE, which in this example is *under*, can be expanded to PLACE - FUNCTION by combining with an argument of the function, in this reading, the function is an object or THING. In (30b), similarly, the simple expression of PATH can be turned into a more complex form, by combining with one of the five functions that map a reference THING or PLACE into a related trajectory, as in "*to the house*" or "*from the tree*". (30c), indicates that, an element in the event category, can be expanded in two ways, either in motion, e.i., GO, or as STAY. The event in the GO function, takes two arguments which are, the THING which is in motion and the PATH, which denotes the direction of motion. On the other hand, in the STAY function, two arguments of THING, which stays and the PLACE, where it is placed. And finally in (30d) a constituent can be expanded through the functions of BE or ORIENT. The BE function, similarly to the stay function in (30c), takes the arguments of THING and PLACE and the ORIENT function, takes the THING argument and the PATH argument which refers to the orientation of an object (Jackendoff, 1987).

In this regard, Jackendoff (1987) proposes a structural definition for the thematic relations of (Gruber, 1965). He argues that the thematic role of *source*, for instance, which is defined as "the object from which motion proceeds," can be structurally defined as the argument of the path-function FROM and *goal*, "the object to which motion proceeds," as the argument of TO.

Later, the Government and Binding framework used Jackendoff's idea about thematic relations in a limited fashion as a syntactic notion, namely, theta-roles. In this new capacity, theta-roles were only assigned to the arguments selected by the predicates. For the correct assignment, two main rules were settled. Firstly, the theta criterion restricts that each argument can only bear only one theta role, and each theta role is assigned to one and only one argument. (Chomsky, 1981)

Moreover, the UTAH explains that Identical thematic relationships between items are represented by identical underlying structural relationships between those items (Baker, 1988).

1.4.3 Proto-Roles

A generalized approach toward the thematic roles was proposed to answer two main issues regarding the thematic roles. Firstly, as Dowty (1989) argues, most of the thematic role lists were unable to make an agreement among the linguists on the number of the roles in a thematic role list. In addition, he argues that there is no consensus on how these roles are allocated to the arguments. Following these two major issues, Dowty argues that the study of the thematic roles has reached an impasse. Thus, Dowty (1991) proposes a generalized framework of thematic roles, instead of a fixed list of them, and in contrast, does not consider them as atomic unanalyzable concepts (Levin, 2019).

In order to shape this frame, Dowty (1991), the conception of thematic roles, takes advantage of the prototype theory of (Rosch and Lloyd, 1978). This framework, in contrast with those of Jackendoff, who considers thematic roles as semantic notions, considers thematic roles as interface notions between syntax and semantics. In this approach, he classifies the roles into two cluster concepts of prototypical roles, e.g., *Proto-Agent* and *Proto-Patient*. These roles are assigned to the verbal arguments based on the lexical entailments the verb imposes on its arguments in the described event (for more about argument structure, see section 1.1).

Dowty considered those features that recur more frequently among the verbs to define these lexical entailments. He proposes that, in general, the agent and the patient of an event can be decomposed into a set of features; for example, he argues that the agent usually has a volitional involvement in the event or state, or they have sentience or perception while taking part in the event; while a patient, undergoes a change of state or could be affected by another participant. Consequently, he proposes a set of features for the proto-roles, which, as he mentions, are not limited to those he mentions, table 1.4 shows a list of these entailments propose by Dowty (1991).

Proto-Agent Properties	Proto-Patient Properties
volitional involvement in the event or state	undergoes change of state
sentience (and/or perception)	incremental theme
causing an event or change of state in another participant	causally affected by another participant
movement (relative to the position of another participant)	stationary relative to movement of another participant
exists independently of the event named by the verb)	does not exist independently of the event, or not at all

Table 1.4: Proto-role properties

As a mechanism of allocation of the proto-roles, Dowty (1991) explains that in predicates with grammatical subject and object, the argument for which the predicate entails the greatest number of Proto-Agent properties will most probably be lexicalized as the subject of the predicate; the argument having the greatest number of Proto-Patient entailments is likely to be lexicalized as the direct object. However, If two arguments of a relation have equal or approximately numbers of entailed Proto-Agent and Proto-Patient properties, then either or both may be lexicalized as the subject or objects.

He also specifies that the proto-roles do not classify the arguments exhaustively or uniquely, or discretely; namely, some arguments might have neither roles, or they can share the same role, or they could qualify partially but not equally for both proto-roles. According to Dowty's perspective, in cases where two or more verbs describe an event but from two different aspects, if the entailments are the same among all arguments, both of them can have a role based on the verb. An example of this could be in verbs such as *buy* and *sell* in 31, in which he argues that the *seller* and the *buyer* have almost the same proto-role features, that is, both of them are taking part in the event in a volitional matter and affecting the other participant; therefore, and they can both be the agent or patient based on the way the event has been looked upon.

- (31) a. John sold the piano to Mary for \$1,000.
 b. Mary bought the piano from John for \$1,000.

He continues that, however, this is not the case for all of the verbs of this type. In cases such as the mental verbs of Croft (1986), the two arguments do not entail the same features as in buy and sell. For instance, verbs such as *like* and *please* in (32) the subject in *like* is entailed to be sentient, though the object is not (the subject has the features of an experiencer) and in *please* the subject causes some emotional reaction or cognitive judgment in the object. In other words, it can be assumed that the mapping Dowty suggests is highly related to the lexical features and the events, and in cases where the same entity can participate on both sides, the mapping might not work well; however, in cases where either the lexical entailment or the event distinguish them the mapping works well.

- (32) a. John likes Marry
 b. Marry pleases John

The understanding provided by Dowty (1991)'s framework has gathered the attention of many scholars interested in experimental linguistic fields since the decomposition of the roles enables the researchers to use the features for examining

them from more different aspects [Kako \(2006\)](#). However, perhaps one of the shortcomings of this approach is that it does not differentiate the degrees of agentivity and retentivity clearly; therefore, the boundaries of the agent and the patient roles could be quite vague in some instances.

1.4.4 Semantic Macroroles

One of the other generalized readings of semantic relations is proposed by [Van Valin and LaPolla \(1997\)](#) as a key component of Role and Reference Grammar (RRG). The [fig. 1.3](#) indicates the categorization RRG draws for the hierarchy of roles. Having scrutinized the graph, it can be observed that the first layer concerns the **Verb-Specific level**, which are the semantic roles assigned based on the description each verb encodes; for instance, for the verb *give*, there should be a verb-specific role of *giver*, a *given* and a *given to*. The second layer is the generalizations over the verb-specific roles, e.g., **thematic roles**; this layer includes labels such as agent, instrument, experiencer, theme, and patient, which are almost the traditional thematic roles. The last semantic layer, which is the generalizations over the thematic roles, is what they consider as the **macro-roles**, which is made of two macro-roles, namely, *Actor* and *Undergoer* ([Van Valin, 2001](#)).

[Van Valin and LaPolla \(1997\)](#) intended to define the roles based on the mechanism of lexical representation of RRG, which is based upon the Aktionsart, e.g., the lexical decompositional features proposed by [Vendler \(1957\)](#), indicated in [table 1.5](#).

Verb Class	Logical Structure
STATE	predicate' (x) or (x, y)
ACTIVITY	do' (x, [predicate' (x) or (x, y)])
ACHIEVEMENT	INGR predicate' (x) or (x, y), or INGR do' (x, [predicate' (x) or (x, y)])
ACCOMPLISHMENT	BECOME predicate' (x) or (x, y), or BECOME do' (x, [predicate' (x) or (x, y)])
ACTIVE ACCOMPLISHMENT	do' (x, [predicate ₁ ' (x, (y))]) BECOME predicate ₂ (z, x)or(y)
CAUSATIVE	α CAUSE β , where α, β are LSs of any type

Table 1.5: Lexical representations for Aktionsart classes

In this framework, the same as those of [Jackendoff \(1972\)](#), the thematic relations are defined based on their position in the logical structures. Considering [table 1.5](#) the state and activity verb classes are able to assign thematic roles, for instance the activity verb *murder*, which its logical structure is represented in [33](#), can assign two thematic roles to its arguments; namely the first argument which is *murderer*

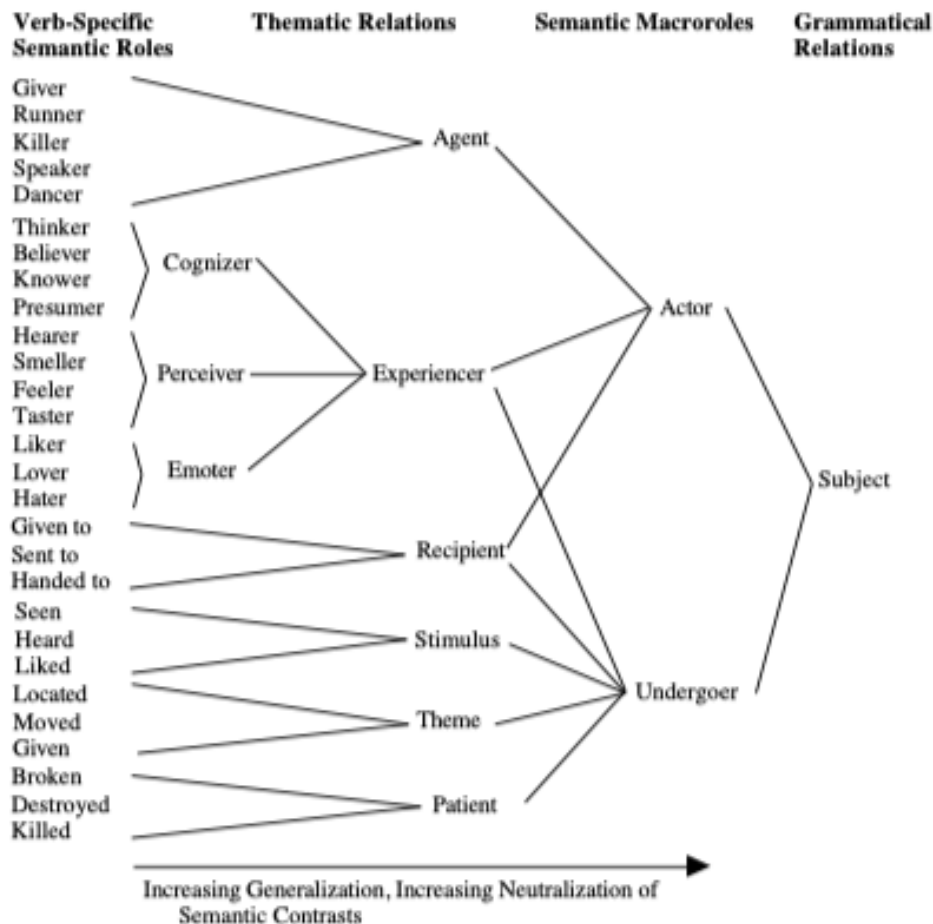


Figure 1.3: Continuum from verb-specific semantic roles to grammatical relations

(subject) as the doer, or agent and the second argument or in this case *dead* (object) as the patient. The thematic relations of all other verb classes are derived compositionally from their constituent state and activity predicates, such as the verb *break* in *A falling tree branch broke the window* and in *Pat accidentally broke the window*, where in the first sentence the agent of the action is not volitional, and *breaker* cannot be defined directly as the agent, the same applies in the second one in which degree of agentivity is lower in comparison with those of *murder*.

To have a better sense of the Aktionsarts and the macro-roles, let us consider the verbs suggested by [Van Valin and LaPolla \(1997\)](#) in (33) for the definitions of the actor and the undergoer.

- (33) a. murder: DO(x,[do (x,)] CAUSE [BECOME dead (y)])
 b. kill: [do (x,)] CAUSE [BECOME dead (y)]
 c. see: see (x,y)
 d. present: [do (x,)] CAUSE [BECOME have (y,z)]

e. *put*: [do (x,)] CAUSE [BECOME be-LOC (y,z)]

The continuum suggested by Pustejovsky and Batiukova (2019) is almost the pivoted version of the hierarchy in fig. 1.3, indicating the standings of the macro-roles and their logical structure along with the thematic roles they can obtain. As Van Valin and LaPolla (1997), argues a prototypical Actor is the subject of a verb that only has an agentive interpretation. As it can be seen in fig. 1.4, it stands on the left side of the vector. Considering the table 1.5, in other words, a prototypical Actor is the subject of the activity verbs, such as those of *murder* which was discussed above, and its logical structure was illustrated in (33). On the other side, a prototypical UNDERGOER as Van Valin and LaPolla (1997) defines the the only argument of the stative semantic predicate which in the above-mentioned verbs it could be the *dead* in *murder* and *kill*, or as the second argument of the stative verbs, such as *see*, or have in *present*, and Be-loc in *put*.

Considering the definitions of the prototypical actor and prototypical Undergoer, let us observe how the continuum in fig. 1.4 accounts for the thematic roles. As the direction of the arrows in the continuum indicate, the highest-ranked actor which is the prototypical actor explained earlier, is positioned at the far left. Moving to the right, the features of the actor, or agentivity, starts to decline, as, in the thematic roles of effector, the volition attribute has reduced; it can become more apparent contrasting the logical structure of the verbs *kill* which has a degree of accident in it, and *murder*, which encodes the will and volition, in 33. Moving further to the middle of the continuum, it can be seen that thematic roles such as experiencer are located almost at the middle of the vector. Thematic roles, which are less agentive, as an example of these roles can be the first argument of the verb *see* that can be termed as experiencer. Considering the status of the Undergoer, as it was mentioned, the highest rank of it is the only argument of the stative semantic predicate such as *murder* and *kill*, moving to the left of the vector, the Undergoer can be the stimulus which is automatically considered as the second argument of verbs such as *see*. Considering the verbs of *put* and *present* in 33, it can be seen that their second arguments, e.g., have, and be-LOC are the only non-actor elements; therefore, they do not have many of the undergoer features, but they can be considered to have the thematic roles of possessed in terms of *present* and location in terms of the *put*.

As it can be remarked, a critical difference between Van Valin and LaPolla (1997)'s macro-roles and Dowty (1991)'s proto-roles is that, in RRG, the semantic roles are associated with primitive semantic predicates into which the verb meaning is decomposed. While for Dowty (1991), the generalized roles are broken down into

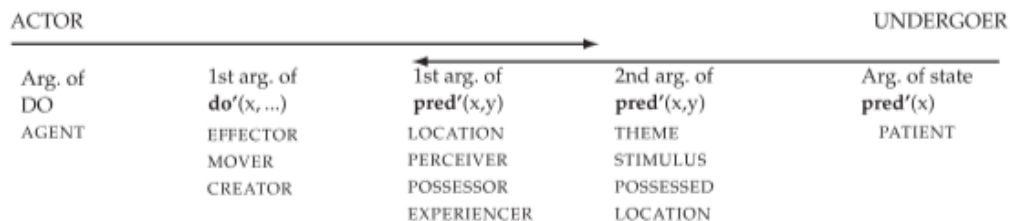


Figure 1.4: Thematic relations continuum in terms of Logical Structure of argument positions

basic semantic features. In addition, the characteristics of the Macro-roles and their correspondence to the thematic are more sophisticated and are able to account for both the more generalized approaches and more specific roles. One intuition that can be obtained from the Macro-role approach of [Van Valin and LaPolla \(1997\)](#), is that in generalized approaches, the thematic roles such as effector and experiencer can be placed under the umbrella term of Actor in general agent-like, and the theme, stimulus, and so on can fit under the cluster of undergoer or patient-like generalized roles.

1.4.5 Thematic roles and Argument Structure Processing

Having reviewed the literature on the thematic roles, a question that needs to be answered for the current research is how each of the reviewed interpretations of the thematic roles is able to reflect and satisfy the needs of the research.

Considering this research which is in search of an answer for the role of the elements of the Persian complex predicates in allocating the thematic roles, based on the reading of the participants. The most important aspect that has to be pointed out is that the roles and the thematic role framework need to truly reflect the understanding or, to be more accurate, the cognitive biases of the participants towards the roles.

As it was mentioned in the first section, verbs in argument structures convey the message as in equations; and the verbal arguments or the participants are like the concepts on different sides of the equation. It was argued that the thematic roles are labels we assign to the arguments based on their role and relationships. Nevertheless, the important question is how these labels are represented and understood in our minds. Do they genuinely exist and are used in our language understanding, or are they only linguistic notions?

One of the principles of the cognitive sciences considers the way human beings categorize the world around them. From this perspective, the study of thematic

roles in cognitive sciences is the study of event-participant categories and the way these categories provide an opportunity to gain insight into the interface between conceptual and semantic knowledge.

Scholars such as [Strickland \(2017\)](#) argue that grammatical systems often represent distinctions that reflect those differences that are rooted in the core knowledge; for instance, the mass and count noun distinctions, which are rooted in the distinction between an object and a substance. From their standpoint, thematic roles have the same characteristics. Considering the example in [34](#), although the verb and event in the instances change, there is a common sense that the role of John and the door is similar in all of them. Suggesting that, regardless of the differences in the approaches towards thematic roles, it can be assumed that the relevance of such thematic roles in creating mappings between semantics and syntax is a common linguistic device across many languages.

- (34) a. John broke the door.
 b. John punched the door.
 c. John painted the door.

[Strickland \(2017\)](#), argues that this distinction in the roles is one of the systematically innate biases in the course of the process of language evolution and is considered a piece of universal core knowledge.

On the other hand, many scholars have argued that thematic roles are, in fact, not that clearly portrayed conceptions. The fact that there is hardly a consensus on a list of them or a fixed definition of the roles also points to the same problem. [Dowty \(1991\)](#), in the introduction of his paper, states that "no concept in modern syntactic and semantic theory which is so often involved in so wide a range of contexts, but on which there is so little agreement as to its nature and definition, as thematic role." A notable example of these controversies could be the instrument role. Regarding the examples from [Rissman and Majid \(2019\)](#), it can be seen that this role contains a wide range of features that contains different types of the argument, and all of them could be interpreted as the instrument.

- (35) a. Janine ate the custard with **a spoon**.
 b. Wanda accidentally drew on her shirt **with a marker**.
 c. Renée applied the lipstick with **her fingertips**.
 d. Carlos carried the milk **in a bucket**.
 e. Anita went to Amsterdam **by train**.
 f. The cue ball hit **the red ball** which sunk the eight ball.

- g. Tyrell used **the steamy room** to practice yoga.
- h. **The bomb blast** destroyed the abandoned factory.
- i. **The program** completed the algorithm in five seconds.

Following these issues, [Rissman and Majid \(2019\)](#) argues that all of the roles could not be tackled in the same way. In their study, by reviewing a range of literature on psycholinguistic and event cognition experiments with adults, children, and infants; typological studies grounded in cross-linguistic data; and studies of emerging sign languages, they argue that the answer to these questions depends on the role. Their study is concentrated on the pair of Agent and Patient, Goal and Recipient, and the instrument. The finding of their study illuminates strong evidence for abstract role categories and a universal bias which provides the ability to draw a distinction between the Agent and Patient roles. For Goals and Recipients, it was found that there is clear evidence for abstraction but mixed evidence as to whether there is a bias to encode Goals and Recipients as part of one or two distinct categories. Ultimately, their results implied that there is no clear evidence for the Instrumental role of either abstraction or universal biases to structure instrumental categories.

Building upon the received wisdom from the findings of [Rissman and Majid \(2019\)](#), it can be argued that the interest of this research does not account for the types or the definitions of the thematic roles while it is exploring the characteristics of the CPr counterparts. Among the reviewed literature on thematic roles, the generalized approaches seem to fit our research question reasonably, following the perspectives of [Dowty \(1991\)](#), who builds his framework of proto-role for two cluster roles of Agent and Patient, based on the proto-typical clusters and lexical entailment. The present study can be based on the differences between these two roles and can examine the role of the CPr counterparts more swiftly, taking the [Dowty \(1991\)](#)'s framework.

1.5 Related Studies

In particular, agent and patient roles are one of the central conceptions of the thematic roles. The reason for this can be traced back to many behavioral studies in linguistics, where there are many pieces of evidence suggesting that the agent and patient roles are represented as abstract categories in our minds. For instance, [Hafri et al. \(2013\)](#) observed that English-speaking adults in their experiment were able to identify and answer questions about the subjects and the objects of transitive sentences in a quite short time; indicating that, during this process, they were

using an abstract schema; other than only verb-specific features, for decoding the events. As reviewed in the previous section, [Rissman and Majid \(2019\)](#) also argues that there is a universal bias in encoding the Agent and Patient roles based on the literature review of the related studies. They explain that these roles guide cognitive and linguistic processing in adults. They explain that these roles are distinctly coded cross-linguistically; in addition, they shape language emergence at an early stage as a part of young infants' conceptual knowledge. As discussed earlier, the framework suggested by [Dowty \(1991\)](#) provides a clustering of the agent and patient roles, enabling the researchers to have a better understanding of these roles or categories based on a set of features.

Considering the evidence that these roles are part of the conceptual knowledge of infants and play an important role in deciphering an event in our everyday language use, an interesting point in complex predicates is how they define these roles. The theories on CPs have mentioned different processes for the allocation of thematic roles; however, studies experimenting with them are rare in the literature.

In the following section, the related studies on complex predicates and thematic roles have been touched on, and in the subsequent section, the research questions are elaborated, and the possible outcomes of the research are discussed.

1.5.1 Studies on Persian Complex Predicates

Complex Predicates in Persian have been the center of many theoretical debates, and they have been studied in different theoretical frameworks. Although their constituting elements have been analyzed from different perspectives according to the theoretical frameworks, there are few experimental, psycholinguistic studies on them.

In addition, a review of the literature indicates that there are a high number of studies on the processes of morphologically complex and compound nouns; however, studies on complex/compound verbs are rare. The reason for this could be due to the fact that morphologically complex verbs are not found in many languages. In addition, most of the studies are focused on the derivational or inflectional affixes rather than on the complex structure of the verbs. [Shabani-Jadidi \(2014\)](#)

[Nilipour \(2000\)](#)'s research on two aphasic patients indicates the disruption of the verbs and grammatical morphemes are more plausible rather than the other linguistic categories. In addition, in the case which had more speaking abilities in the majority of their utterances, the lexical verb was omitted, or it was replaced by the *ast*, 'is.' Nilipour concludes that the manifestation of disruption in Persian verbs can be deletion of the verb; substitution of the infinitive for the proper inflected

form of the verb; substitution of a filler verb *budan*, "to be" for the appropriate lexical verb; deletion and/or substitution of the inflectional morphemes and/or verb prefixes; broken-off forms producing partial words.

Later, Nilipour and Raghibdoust (2001) conducted their research on seven aphasic patients; in this research which was intended to discover the general characteristics of the agrammatic language in Persian, it was observed that, almost in all of the cases, there is an interruption in the verb, deletion of free grammatical morphemes, the substitution of bound grammatical morphemes, and reliance in nouns rather than verbs. An example of the substitution of the verb can be seen in (36) where the verb *fod*, 'become' has been interrupted.

- (36) yedafe pare -od
 suddenly tore was.3SG
 "It suddenly tore off."

In addition to the studies on aphasic patients, Shabani-Jadidi (2014) also studies how Persian compound verbs are being processed in the mental lexicon and how they are stored, organized, and accessed. Their conclusion indicates that it is possible to use the compound verb as a noun, namely, the nominalization of compound verbs is possible; next, they argue that there is a non-referentiality of the nominal constituent in the compound verb, i.e., the nominal constituent cannot be followed by a pronoun that refers to it. This research sheds light on several important factors in theoretical issues. Firstly it argues that compound verbs in Persian are decomposed into their constituents at the early stages of processing. In addition, it argues that the compound verbs at the early stages of processing decomposition are based on purely orthographic similarity, and lastly, although both transparent and opaque compound constituents were facilitated while processing, transparency impacts processing in the early stages of processing.

The results of the Shabani-Jadidi (2014)'s experiments, and Nilipour (2000), and Nilipour and Raghibdoust (2001) studies on aphasic patients indicate that CPs are mentally processed in a decompositional way; however, the results of Shabani-Jadidi support a lexical approach to compound verbs in Persian. In other words, Shabani-Jadidi argues that complex verb constituents are accessed at the prelexical processing stage, and syntactic calculations are done at later stages of processing. This conclusion indicates that the non-verbal elements, as Goldberg et al. (2003) argues, work as the *hosts* in CPrs, and the light verb, which takes the syntactic features, is later incorporated.

1.5.2 Studies on Thematic Roles

The other aspect of the research that deals with the thematic roles and, more particularly, the proto-roles of Dowty (1991) has been tested and used in the works of many scholars of psycholinguists, cognitive sciences, and computational modeling. One of the most important assets of Dowty's framework is that he offers a decompositional perspective of the thematic roles, enabling researchers to study roles not as atomic concepts but based on their features. Having reviewed the literature, it can be said that there are two main approaches to understanding Dowty's proto-role framework.

The experiments of McRae et al. (1997) on the participants' rankings of role/filler featural similarity for Agent and Patient roles suggests a verb-specific reading of thematic roles. In McRae et al. (1997)'s perspective, thematic roles are considered as slot and filler mechanisms. Namely, for a verb such as *accuse*, there are two slots that need to be filled with relative elements to the verbs that are the person who accuses (accuser) and the person who is being accused. In this regard, he performs several experiments, examining the characteristics of these roles based on the verb-specific features. In some senses, this account of the thematic roles reflects those of Dowty (1991), but with the difference that, in this regard, the roles are considered as concepts highly dependent on verbs.

McRae et al. (1997), begins his experiments by deriving a number of features from the participants for 20 transitive verbs which can only have an animate agent and patients, such as *teach*, *rescue*, *hire*, and so on. In the next step of his research, he measures the typicality of the feature by asking the participant questions such as *how likely is it for a monster to frighten someone?*. In the next step, by asking the participants to rate the importance of the noun and its features, they measured the importance of features compared to the nominals. Their results indicate that the features' importance is parallel to those of the nominals. In the final part of their experiment, a self-paced reading task was conducted by adding irrelevant features to the nominal element of the constructions. As a result, it was observed that there is a significant effect of the verb-specific features on the participants' comprehension, suggesting the effect of the thematic role features in our everyday language use.

As a result of this research, it has been concluded that the thematic roles might best be viewed as verb-specific and that this conceptual knowledge is computed and used in online language processing. Scholars in different languages have replicated this reading of the thematic roles. Lebani et al. (2015), investigates a set of 20 Italian verbs following the experiments of McRae.

The verb-specific nature of McRae’s study also prepares the ground for the studies in computational modeling studies and feature-based experiments. For instance, [Chang \(2004\)](#) presents a verb-specific model that learns the association between form relations and meaning relations; however, the model has limited a set of verbs and does not generalize beyond the scope of individual verbs. In a more recent study, [Lebani and Lenci \(2018\)](#) presents a verb-specific model inference by exploiting corpus-based distributional data and proposes a new approach to the representation of thematic roles.

From the perspective of scholars such as [Kako \(2006\)](#) the thematic roles are considered as a model which has three sides; one side of the model is the grammatical position, the other side is the lexical and semantic features, and the last side considers the thematic role. Taking this modeling of the thematic roles in mind, [Kako \(2006\)](#) investigate the relation between the grammatical role of the sentences and the proto-roles in the absence of any verb-specific information. In this experiment, Kako uses a questionnaire made of a set of questions about the proto-features of the proto-role hypothesis. A summary of the questions and their corresponding features can be seen in table 1.6.

Proto-Agent properties How likely is it that the [Subject]...	Labels
Chose to be involved in [verbing]?	Chose
Was aware of being involved in [verbing]?	Aware
Caused a change in the [object]?	Caused Change
Caused the [object] to do something?	Caused do
Moved?	Moved
Existed before [verbing] took place?	Existed
Proto-Patient properties How likely is it that the [object]...	
Was changed in some way as a result of [verbing]?	Changed
Was created as a result of [verbing]?	Created
Was stationary?	Stationary

Table 1.6: Kako’s (2005) proto-features

In this experiment, participants were asked to answer the above questions about the objects and the subject in several sentences. Kako surveys structures with meaningful and real verbs in the first two experiments. As it was expected based on the theory of [Dowty \(1991\)](#), the subjects showed more agentive-like characteristics, and the objects entailed more patient-like features. In the second phase of the study, Kako substitutes the real verbs with nonce content words such as (*"The rom mecked the zarg"*). The results of this phase of the study were the same as those of the first two, highlighting that the roles are highly related to the grammatical positions. In the next step of their analysis, they confirm that there is no effect of the subject and object differences in terms of grammatically irrelevant properties,

such as the effect of *"being liquid"* on the subject or object features. Furthermore, in the final phase of their study, Kako uses verbs that do not permit objects, such as "erupt" and "fall," in sentences with an object; namely, the verbs were inserted in constructions they were not able to build. The results of this final step also indicate that the Subject tends to be more agent-like while the object is more patient-like. The results of his experiment shed light on the fact that the reading of subject and object is highly related to the concept of agent and patient in most cases; namely, entities in the subject position have a higher tendency to be a proto-agent than those of the object are more proto-patient like. In addition, the results of the last experiment demonstrate that the participants, when inferring the thematic roles link the grammatical position and the semantic features. In a general deduction, it can be stated that the trend observed in all of the conditions was almost the same. To be specific, the proto-agent features were more observed in the subject position, and the proto-patient was mostly observed in the object position.

Following Kako's study, [Reisinger et al. \(2015\)](#) replicates the survey on a larger scale of participants and with some changes. In this research, the authors followed the same methodology in terms of the experiments. Their survey was made of several questions based on the proto-features of [Dowty \(1991\)](#). This research, the same as those of [Kako \(2006\)](#) is made of nonce-based and corpus-based instances. The result of their study also confirms the finding of [Kako \(2006\)](#); namely, there is a higher tendency to have proto-agent roles in the subject and to have proto-patient in those of the object. In addition, this research indicates the possibility of decomposing the thematic role into smaller fragments that can be defined through annotation tasks. This study also introduces a protocol for this annotation called Semantic Proto-role Protocol (SPR). In a follow-up study, [White et al. \(2016\)](#) suggests some revisions to the protocol and introduces SPR2. As [White et al. \(2016\)](#) explains, some of the feature questions which were redundant or sources of error were revised. They summarize that the redundancies included the feature of *stationary* being essentially the negation of *change of location* and *predicate changed argument* being almost identical to *change of state*. In addition, *the property exists as physical* was dropped because it is a purely referential property and non-relational property of the argument; thus, it is redundant with their decompositional word sense protocol. Moreover, they removed *The location of the event* and *physical contact* properties because of lower inter-annotator agreement and high within-annotator response variance in the first version. On the other hand, they added three properties of *benefactives*, *partitives*, and *incremental themes*. They argued that the Benefactive and partitive arguments often appear in special morphosyntactic characteristics. As an instance for the *Benefactive*, in English, in most cases, they are used in

double object constructions such as *John bought her a book*. Moreover, the *partitive* accounts for the partial affectedness, which in many languages is marked with morphological cases. The third new property, *change of state continuous*, is a reformulated and simpler version of Dowty’s (1991), *incremental theme* which was not included in [Reisinger et al. \(2015\)](#). The following table 1.7 indicates the final SPR2 protocol and the role properties and the related questions for the annotation.

Role property	How likely or unlikely is it that...
instigation	ARG caused the PRED to happen?
volition	ARG chose to be involved in the PRED?
awareness	ARG was/were aware of being involved in the PRED?
sentient	ARG was/were sentient?
change of location	ARG changed location during the PRED?
existed before	ARG existed before the PRED began?
existed during	ARG existed during the PRED?
existed after	ARG existed after the PRED stopped?
change of possession	ARG changed possession during the PRED?
change of state	ARG was/were changed during or by the end of the PRED?
was used	ARG was/was used in carrying out the PRED?
was for benefit	PRED happened for the benefit of ARG?
partitive	Only a part or portion of ARG was involved in the PRED?
change of state continuous	The change in ARG happened throughout the PRED?

Table 1.7: Semantic Proto-role Protocol 2

The protocols used in this experiment were a part of the Decomp project, and the complete set of verbs and the dataset of the annotations derived from both of these studies can be accessed through the Decomp website documented in [White et al. \(2019\)](#).

The insights yielded from the thematic roles model suggested by [Kako \(2006\)](#) have inspired more sophisticated models based on the semantic properties and the syntactic positions, regardless of the verb-specific features. For example, [Alishahi and Stevenson \(2010\)](#) propose a probabilistic usage-based model of semantic role learning. Their model can acquire associations between the semantic properties of the arguments of an event and the syntactic positions that the arguments appear in. In other words, the model discovers the connection between the syntactic patterns of an utterance and their semantic properties; as a result, the common properties are gathered in a profile, and based on the connections, they represent the roles associated with the position. The intriguing outcome of this model considers the fact that it delivers us an insight into how the intuition regarding the general roles such Agent and Patient are shaped.

1.6 The Current Study

The perception formed by the review of the literature demonstrates that the notion of the argument structures and the thematic roles are key notions in understanding language. A cognitive interpretation of the thematic roles indicates that, in particular, the Agent and Patient roles are universal biases we use for understanding the language.

Taking the importance of the thematic roles into consideration. An intriguing case study can be the Persian Complex Predicates. As the literature review illuminates the diverse interpretations of the CPRs, it can be argued that there are three main theoretical analyses regarding the components of these constructions.

Firstly the Lexicalist account of [Karimi-Doostan \(1997, 2005\)](#) on the CPRs illustrates that the meaning of the constructions is forming the event structure. In this respect, the counterpart which encodes the meaning of the whole construction is the element that denotes the thematic roles. In their perspective, the LVs are semantically bleached and are not able to affect the event structure. Regarding this perspective, the whole emphasis is on the NV element in terms of the thematic role.

The other account of the CPRs, which is rooted in the constructionalist approach of [Folli et al. \(2005\)](#); [Megerdooonian \(2001\)](#) and in contrast with the lexicalists, argues that the LV elements are the counterparts that are responsible for the formation of the structure, and the other elements are selected based on their characteristics. In this point of view, the NVs, the same as those in the lexicalist approach, are viewed as the elements encoding the event and event type, while the LVs perform the thematic role allocation, coding duration, and indicating eventiveness.

In contrast with those mentioned above, the last reading of the CPRs we reviewed here argues for a posterior analysis of the CPRs, arguing that although these constructions have linguistically independent elements, they cannot be analyzed in isolation. Therefore, a construction-based approach of [Samvelian and Faghiri \(2013b\)](#) and [Goldberg \(1996\)](#) propose that the CPRs need to be classified by the contexts, that is, the meaning of them as a whole, and based on the constructions they appear in. This understanding of the CPRs considers the thematic role allocation as a consequence of the interaction of the CPRs.

Bearing in mind that thematic roles can be seen as cognitive schemata affecting our language understanding, this research is trying to solve how Persian CPRs select the thematic roles and which of the interpretations of it can better explain their correspondence.

In order to tackle this question, we are using the decompositional approach

toward the proto-role hypothesis of the Dowty (1991). In other words, in a decompositional approach such as the studies of White et al. (2016) and Kako (2006), a three-way model is considered. Namely, the semantic properties, syntactic position, and the proto-roles are principles of the model. Since we are investigating the thematic roles, the semantic properties and the syntactic position are set to be controlled; namely, entities containing the semantic features of the roles are used to fill in the blank in the subject position of the CPr constructions.

To tackle the question of the effect of the constituting elements of the CPrs, the participants' judgments can be seen as the input of a model, and the light verb and the non-verbal part are set to be the variables affecting the judgment. Our probabilistic model is set to reveal the probability of having different proto-roles due to the change in the variables, demonstrating how each of them can affect the judgments. The state-of-the-art of the current research is that it is not accounting for specific instances to reject or confirm the aforementioned studies, while it considers a relatively high number of the instances and scrutinizes the trends and points of dissimilarity in these constructions.

The results obtained from the model can shed light on how the thematic roles in Complex Predicates are allocated and which of the theories on the roles of the constituents can better explain their correspondence. Moreover, it can yield insight on how different the components of the CPrs can affect the readings of the thematic roles and in which conditions their significance change.

Chapter 2

Research Methodology

The current research intends to answer the key question following the literature on Complex Predicates and the notion of thematic roles. As discussed in the literature review, the role of the CPr elements in assigning the thematic role to the arguments of CPr constructions is a matter of controversy among scholars. This research intends to provide a model of the constituting elements of the CPrs, based on the participants' judgment for the selection of the proto-roles, to shed light on the trends of the changes among the light verbs and in different conditions of the non-verbal elements. Since this research is focused on the subject position of the constructions, following the experiments of [Kako \(2006\)](#) , [Reisinger et al. \(2015\)](#) and [White et al. \(2016\)](#), it is expected to observe more tendency toward the selection of agent roles; however, the literature on the CPrs indicates that the role could change with the change in the constituting elements of the sentences.

2.1 Methods

In order to gather data about the research question, an online questionnaire was devised containing CPr constructions, asking the participants to make judgments according to their linguistic intuitions. The following sections explain how different parts of the questionnaire were selected, who the participants were, what procedure they had to undergo for filling out the questionnaire, and how data were extracted and analyzed.

2.1.1 Material

In this experiment, a fill-in-the-blank questionnaire was designed, which contains CPr constructions formed by eight frequently-used LVs in today's Persian, and are able to create constructions according to our criteria. All optional elements were

removed to control the different constituting elements of the sentences. Besides, to reduce the possible role of the meaning of the direct objects, they were replaced by either something, someone, or somewhere, based on the meaning of the verbs.

Since this research intends to investigate the subject position of the constructions, the subject position is blank, and the participants are asked to fill in the blanks with two choices which are entities representing the proto-roles. In the following sections, I explain more about the selection of light verbs, and non-verbals and how the proto-role entities were derived from the dataset of [White et al. \(2016\)](#).

Light Verb Selection

In order to filter out and select a number of LVs, several filtering options were applied. In the first step, the list of the identified LVs based on their semantic categorization proposed by [Karimi-dustan and Eshaghi \(2019\)](#) was used as the preliminary list of the verbs. Afterward, the LVs were searched in [Seraji et al. \(2016\)](#)'s corpus of Universal Dependencies for Persian.

As [fig. 2.1](#) indicates, the frequency of *kardan*, 'to do,' is significantly higher than the rest of the studied LVs. On the other hand, several of the cases, e.g., *gaftan*, 'to turn,' *jaftan*, 'to obtain', were not even found in the corpus. Therefore, the LVs which were placed in the first quartile were removed. In addition, LVs, which are significantly frequently used with their heavy meanings such as *didan*, 'to see,' and *residan*, 'to arrive,' were omitted as well.

The rest of the LVs were searched in the dataset of PersPred by [Samvelian and Faghiri \(2013a\)](#), and those which were able to form both transitive and intransitive constructions were derived along with their NV element. As a result, eight LVs that met our criteria were chosen. The list of the selected verbs and their semantic class can be seen in [table 2.1](#).

Verb Class	Light Verb
State Verbs	daftan 'to have'
Do Verbs	kardan 'to do'
Verbs of Creation and Transformation	ʃodan 'to become'
Verbs of Existing Force	keʃidan 'to pull'
Verbs of contact by impact	zadan 'to hit'
Verbs of change of possession	dadan 'to give'
Verbs of Obtaining	gereftan 'to get'
Motion Verbs	ʔa:vardan 'to bring'

Table 2.1: Selected Light Verbs and their Semantic Classification

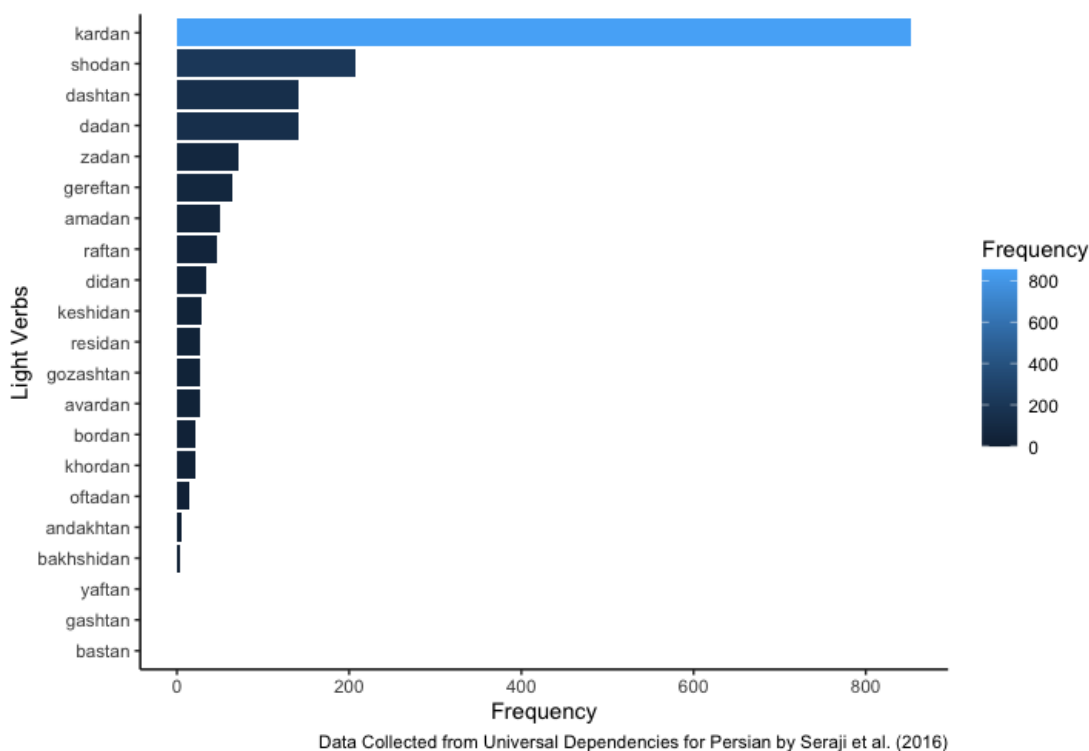


Figure 2.1: the frequency of the LVs in UDP corpus

Non-verbal elements selection

In order to grasp the role of the non-verbal elements, a number of nonce NVs were used. Our hypothesis is that if the NVs play a significant role in the judgment on proto-roles, a significant change in the judgments must be observed in their absence. Consequently, the constructions in the survey contain four meaningful and four nonce NVs for each of the light verbs. To select the meaningful NVs, the criterion suggested by Karimi-Doostan (2005) has been used. He argues that the valency of the CPRs is set according to the argument structure imposed by the NV elements. Two of the selected NVs were able to form transitive, and two forming intransitive construction. For deriving the NVs, the PersPred database of Samvelian and Faghiri (2013a) was used as well. This database allows the application of the filters based on the number of arguments; as a result, for each of the light verbs, the filters of N0 N1 NV LV and N0 NV LV were applied and based on the filterings, NVs were derived. It should be noted that, of the light verb *fodan*, 'to become' there was only one instance in the transitive form in the database; therefore, one NV was added from the observations of the researchers.

In addition, the nonce NVs were used, with the same distribution as those of the meaningful ones for each of the LVs. In the production of the nonce NVs, to follow the phonetic structure of the words in Persian, several meaningful NV elements

were selected, and their phonemes were changed. Several speakers of the language were asked to read aloud the nonce words, and they reported that, although these elements are nonce, they faced no difficulty in reading them. A complete set of the NVs and their LVs can be seen in appendix A.

Filler entity Selection

The next step in forming the questionnaire was related to the selection of proto-role entities encoding the lexical features of the proto-agent and the proto-patient roles. For this stage, the previously studied and annotated dataset of White et al. (2016) was used. To categorize the investigated phrases, the average of the proto-features was calculated for each of the proto-roles. The raw dataset includes data obtained from three semantic proto-role protocols of SPR1, SPR2.1, and SPR2, including 3482 argument phrases. However, not all of the investigated data were applicable. To clean the data, firstly, the not-applicable data were filtered out, and the argument phrases tested through SPR2 and SPR2.1 were selected. Next, the mean of each proto-feature for each argument phrase was calculated. Afterward, the data which were incomplete were excluded from the research. After preparing the clean data, the proto-features were categorized based on their proto-roles. That is the proto-features of 'instigation', 'volition', 'awareness', 'sentient', 'change of location', and 'existed before' were categorized as the features of *proto-agent* and, those of 'existed during', 'existed after', 'change of state', 'change of state continuous', 'was used', 'was for benefit', 'partitive', were grouped as *proto-patient*. In table 2.2, an example of the ratings of the proto-features has been indicated.

Following Dowty (1991), the average rating of proto-agents was subtracted from those of proto-patient. Consequently, an index was devised where the higher negative ratings indicated the higher quality of Proto-Patient roles and the higher positive values represent higher proto-agent features. The distribution of the entities based on this index forms the clustering of proto-roles which their distribution in our data can be seen in fig. 2.2.

Since this research uses these entities as the representation of proto-roles, those with higher qualities of the entailments seem to be a better fit for the study; therefore, proto-roles with the highest ratings in the index regardless of outliers were chosen. The distribution of data suggests, argument phrases with an index between the minimum and the first quartile, e.g., -3.36 and -1.36 as a proto-patient entity, and those between the third quartile and the maximum, e.g., 0.060 and 1.140 , for proto-agent entities. As table 2.3 indicates, the entities such as 'chris and ben,' 'himself,' and 'one of robinson's fans', which almost all of referring to animate

Arg. Phrase	proto-features	Ratings
a marvelous job	Awareness	1
	Change of Location	1
	Change of Possesion	1
	Change of State	4.5
	Change of State Continues	4.5
	Existed After	3
	Existed Before	1
	Existed During	4.5
	Instigation	1
	Partitive	4
	Sentient	1
	Vollition	1
	Was for benefit	5
	Was used	5
	Average of the proto-agent features	
Average of the proto-patient features		4.35

Table 2.2: An example of the proto-features and their ratings from [White et al. \(2016\)](#)

entities, are forming the proto-agent cluster, and phrases such as 'more refineries,' 'headlines,' 'these changes' and 'a marvelous job' form the proto-patient cluster, which are mostly inanimate entities and refer to places or conditions and situations.

To use these entities in Persian, a literal translation does not seem to suffice, because phrases such as the 'the pew researchers' might sound familiar to people of the US or people who are following the news. However, the average speakers of Persian do not seem to be familiar with this research institute; consequently, to prepare more familiar choices for the target participants, we implemented four strategies; in cases such as 'the pew researchers' the head of each argument phrase was derived and used as the choice. In cases such as 'one of robinson's tearful supporters,' the dependent phrase was substituted by a more familiar entity such as the name of a football club. In addition, when the entity is referring to a place, such as 'refineries', the entity is substituted by another place that is more frequently used in the everyday use of language, such as 'universities,' and in cases of the proper nouns, Persian names were replaced. [table 2.3](#), indicate the final set of the proto-role entities used in the questionnaire and their original form in the dataset. A complete list of the questions and their entity filler can be found in [appendix B](#).

It is worth mentioning that perhaps one of the most important limitations of this research is rooted in the way it has been dealt with the proto-role entities. As [Grimm \(2006\)](#) also notes, the Agentivity/Patientivity of a fully interpreted phrase is

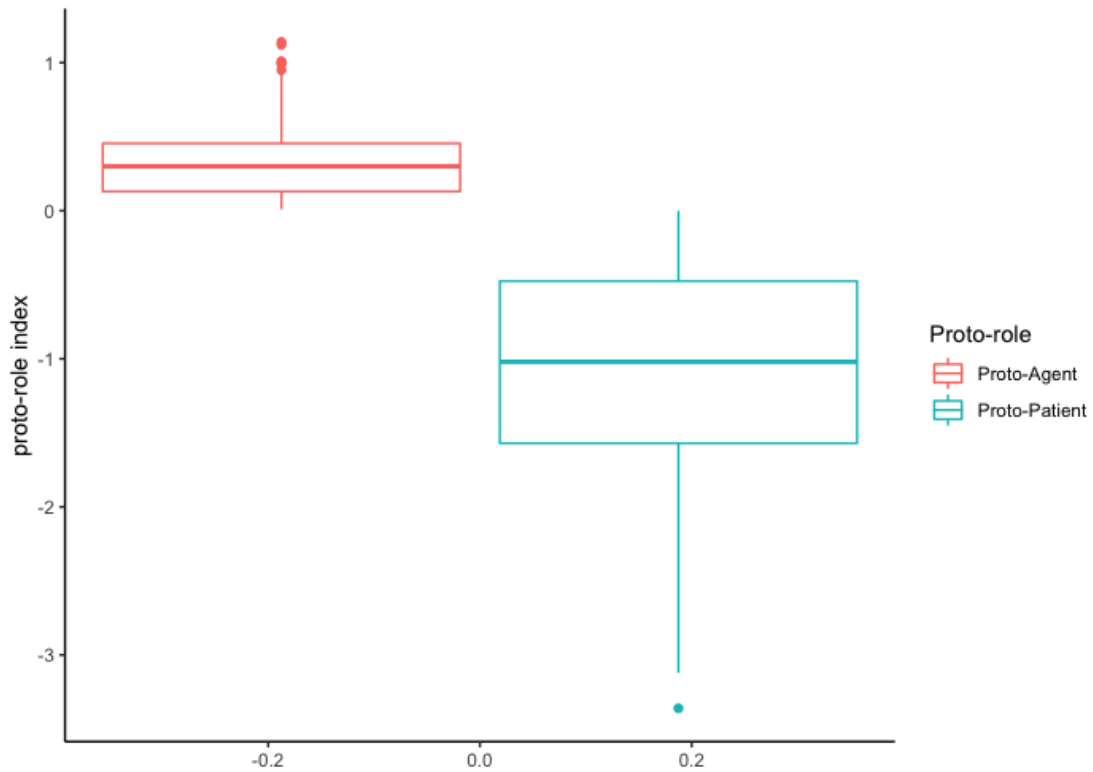


Figure 2.2: Distribution of Proto-role Entities

not the same as the Agentivity/Patientivity inherent in the lexical semantics of the predicate. And since the entities were derived regardless of the verbs co-occurring with them in the experiment, the selected entities mostly encode and carry the lexical semantics of the proto-roles.

2.1.2 Participants

Participants of this research were all native speakers of Iranian Persian (Farsi), mostly residing in Iran, with an average age of 35.30 years. According to the power analysis conducted on the model, with a number of 70 participants, the power for most of the questions regarding their LVs and NVs will reach 75%. However, due to the lack of resources and limitation in time, a sum of 62 participants took part in this research.

2.1.3 Procedure

The questionnaire was designed on the Qualtrics platform, and the participants were invited to participate via the link provided by the website. And mostly from family and friend group chats on social media apps.

At the beginning of the survey, the privacy and terms and conditions of data

Arg. Phrases in White et al. (2016)	Phrases in the questionnaire	Proto-roles
chris and ben Paula one of robinson’s tearful supporters Pew Researchers sumni clerics in the town this dealer the iranians the shattered communist forces	Masoud and Reza Kimea One of Machester United’s Supporters Researchers Teachers The salesperson The Iranians Communists	proto-agent
a marvelous job these changes more refineries the task of guiding our youth down the path of righteousness a kid meal with a milk the intention of the efforts and actions documented in my earlier update	A marvelous Job These changes Universities Our tasks A kid meal our believes	proto-patient

Table 2.3: Proto-role entities used in the questionnaire

processing, with the contact information of the researchers, were provided. After receiving their consent, several demographic questions were asked. Next, participants had to answer seven training questions with validation. The first training question considers the fact that both of the options can be chosen; therefore, an example with a meaningful verb was provided, which can be grammatical and meaningful with both of the proto-roles. The validation of the question can only be passed if both of the options are selected. The next training question considers the nonce NV elements; participants were explained that these complex verbs are from some dialects that perhaps they have not heard of. Later a nonce LV combination which has not been used as the research item with to LVs of *kardan*, 'to do,' and *fodan*, 'to become' have been used. To draw the attention of the subjects and fully introduce the form of the nonce element, they had to answer two conjugation questions regarding the singularity and plurality of subject-verb agreement. Next, they had to answer four questions similar to those used as the research items in the survey; the first question contains a nonce but in context, and the context of the questions reduces gradually as it goes further, as in the last training question subjects face the exact type of question they will face in the questionnaire.

After the training section, the participants had to answer the research items. Since 64 items might negatively impact the attention of the participants, the items were divided into two groups, each of them containing the same type of items, namely eight LVs with nonce and meaningful NVs, in transitive and intransitive contexts. As a result, each subject had to answer 32 items. However, in the end, the two groups were set to receive the same number of responses. It is also worth mentioning that the ordering of the presence of the questions and the filler choices were set to be random for each participant. A screenshot of the questions in the Qualtrcis platform can be seen in 2.3.

Based on the estimate of the survey platform, answering the whole questionnaire

_____ امتحان گرفتند.

<input type="checkbox"/>	محققان
<input type="checkbox"/>	باورهامون

_____ بوماک داشتند.

<input type="checkbox"/>	معلمین
<input type="checkbox"/>	دانشگاهها

_____ درد می کشیدند.

<input type="checkbox"/>	محققان
<input type="checkbox"/>	دانشگاهها

Figure 2.3: screenshot of the questionnaire

would take approximately 10 minutes. The recorded data of the current research also indicates it took an average of 15 minutes to complete the questionnaire.

2.2 Data Analysis

Following our research question on how the non-verbal and the light verb elements can affect the choice of participants on the thematic role of CPR constructions, this experiment intends to shed light on the effect of these elements according to the differences in their contexts and conditions, namely, tackles the questions on how by changing the light verb the decision can be affected, how the non-verbal can affect reading, and in which cases their change is significantly effective. Therefore, by using different frequently used light verbs, it has been tried to observe the changes according to the light verbs. By using the nonce in the non-verbal elements, it has been attempted to observe the role of the non-verbal elements.

To analyze the obtained data and the interactions of these elements, a three-

way generalized linear mixed-effect model was fit to the data in R programming language, using `glmer` from `LM4` package. This model, which is based on the probability assessment, indicates the role of each of LV and NV elements, e.g., predictors, on the dependent variable of this research which is the participants' judgment. Consequently, the model predicts in which contexts the predictors are more significantly effective and where their interactions are.

In the following sections, the overall results obtained from the experiment and the formulation of the model are described.

2.2.1 Data

As a result of this experiment, 3968 observations (judgments) were gathered from 62 participants. The judgments are distributed equally among the light verbs and the non-verbals. Namely, participants answered 248 questions for each light verb element. A cursory observation of the table 2.4 suggests that the correlation mentioned Dowty (1991) on the relationship between the proto-roles and the grammatical positions, that is, the proto-agent role is more likely to be in the subject position, is also replicated in our study as in the studies of Kako (2006), and White et al. (2016).

Light Verbs	non-verbal type	Proto-roles		
		Proto-Agent	Proto-Patient	Both
ʔa:vardan 'to bring'	nonce	31.85	6.45	11.69
	meaningful	41.53	2.01	6.45
dadān 'to give'	nonce	27.01	7.66	15.32
	meaningful	32.25	12.90	4.83
daftān, 'to have'	nonce	29.43	7.25	13.30
	meaningful	17.33	25.40	7.25
gereftān 'to get'	nonce	36.29	4.03	9.67
	meaningful	48.38	0.0	1.61
kardān, 'to do'	nonce	31.85	5.64	12.50
	meaningful	40.32	2.82	6.85
kefīdān 'to pull'	nonce	35.88	4.03	10.08
	meaningful	45.16	1.20	3.62
fodān 'to become'	nonce	33.06	8.87	8.06
	meaningful	22.58	21.77	5.64
zadān 'to hit'	nonce	37.90	3.62	8.46
	meaningful	34.67	12.09	3.22

Table 2.4: Percentage of the Judgments for the proto-roles based on the Light Verb and the Non-Verbal

Having scrutinized the data, it can be observed that, in all of the instances, the majority of the judgments are dedicated to the proto-agent entities; however, it can be seen that the degree of correlation is not the same among all of the variables. Regardless of the effect of the other variables, it can be seen that light verbs of *gerefтан*, 'to get,' and *kefidan*, 'to pull,' subsequently have the highest percentages of judgments on the proto-agents; while those of *daftan*, 'to have,' and *fodan*, 'to become,' possess the lowest among the reported data.

Nevertheless, it is still unclear whether the reason for these differences is rooted in the effect of the light verb or those of the non-verbal element, or both of the variables; therefore, to better understand the data, a generalized linear fixed-effect model has been used. In the following section, the model and its possible interpretations are explained.

2.2.2 Regression Model

Linear Mixed-effect Models have been widely used as one of the sophisticated data analysis methods in psycholinguistic and cognitive studies because, in most of these studies, the data contains crossed random effects, and in many cases, it is not normally distributed. More discussion on their superiority to older models can be found in studies of Jaeger (2008), and Baayen et al. (2008).

Due to the nature of this study, e.g., the interactions between the variables, the binary distribution of the data, and the presence of the random effects, namely, judgments of the participants, a linear mixed-effect model have been fit to the data to provide more insight into the research questions.

Preparation for modeling

The raw data retrieved from the Qualtrics platform were filtered only to contain the completed surveys, and it was exported using the numeric data for the choices. In addition, the values were split into columns, and those which were seen but not chosen were marked as 0.

To clean the data and prepare the data frame for the analysis, the open-source program of R Studio Version 1.4.1717 was used. The obtained results from the questionnaire were, reformated from wide to long, using `pivot_longer` in Wickham and Girlich (2022)'s `tidyr` package. To be more specific, instead of having each participant's judgments in a row, the data pivoted to each instance along with the participants' ID, question number, type of non-verbal (NV), type of light verb (LV), type of the proto-role, and the judgment per row.

The model

After cleaning and preparing data for the analysis, a linear fixed-effect model was used. According to the type of the data, which contains acceptable/unacceptable judgments from the participants, namely categorical data, the generalized linear fixed-effect model was used.

As discussed in the previous sections, the main question of this experiment is to discover the role of the constituting elements of the complex predicates; hence, the non-verbal, light-verb, and the proto-roles were considered as the independent variables, or to be more precise in this linear models, as fixed effects. One of the sophisticated advantages of the linear models is that they also account for some of the variables that might affect the experiment while they are not controlled, e.g., random effect. In this research, random effects are set to be each of the items because it would be impossible to have all of the instances of the NV LV incorporations and the participants since; although the mother tongue of all of them is Iranian Persian (Farsi), they are from different age groups and some of them living in different parts of the the country; therefore, their judgments might be affected by those backgrounds. In addition, since one of the possible answers to our research question is the interplay of both elements, we are using a regression model accounting for the interaction of the variables, where both NVs and LVs are the critical elements.

In order to find the fittest model among the variables, we followed (? , p. 268)'s backward step-wise testing of the different interactions of the fixed effects with the proto-roles. In order to perform this comparison, we developed three models. One of the models considers the interaction of the NV and LV and the proto-roles on the judgments. The other two models, were smaller versions, one of them made of the interaction of the NV and the proto-roles; and the other one contains the interaction of the LVs and the proto-roles. The comparison of the three-way model and the two-way models was drawn using ANOVA and its results are summarized in table 2.5.

Model	AIC	BIC	<i>P value</i>
NV \times proto-role	3747.4	3784.3	
LV \times proto-role	3557.8	3656.2	<0.001
(LV \times NV) \times proto-role	3427.2	3611.8	<0.001
LV \times proto-role	3557.8	3656.2	
(LV \times NV) \times proto-role	3427.2	3611.8	<0.001

Table 2.5: Comparison of different models, according to the hypotheses

The comparison indicates that with the presence of all of the variables the model has the highest significance; in addition, the comparison of the *AIC* and *BIC* of the models suggest that the three-way interaction model can be the best fit since it has the lowest among the other models.

A three-way interaction logistic regression model enables us to investigate and compare the role of the fixed effects by setting one of the conditions as the reference point and comparing the rest of the data according to that. Therefore, all of the data are dummy coded, which means that the data are coded in a binary mode of 0 and 1. Those coded as 0 are the reference points of the model. A mathematical representation of a three-way linear mixed-effect model is described in 2.1.

$$y_j = a + a_p + a_i + b_{nv}x_{nv} + b_{lv}x_{lv} + b_{pr}x_{pr} + b_i x_{nv}x_{lv}x_{pr} + e_i \quad (2.1)$$

The left side of the equation includes the dependent variable marked by y_j , which indicates the participants' judgment. On the other side, the a in the equation indicates the intercept or the reference point, which is where all of the variables are those coded as 0. The a_p and a_i stand for the random effects of the model, namely the effect of the variety of the participants and each of the instances. b_{nv} and b_{lv} and b_{pr} successively account for the slopes of the non-verbal, light verb, and the proto-role fixed effects. The x_{nv} and x_{lv} and x_{pr} are for the specific x values of the variables. The chunk of $b_i x_{nv}x_{lv}x_{pr}$, considers the interactions of the variables, and e_i stands for the residuals of the model.

In order to run the model and perform the calculations on the data frame, the `glmer` function from [Bates et al. \(2015\)](#)'s `lme4` package was used. The above-mentioned equation was formulated in the R environment:

```
mmmmodel <- glmer(Judgment ~
  NV*LV*protoroles+
  (1|ID) + (1|question),
  data = test1, family = binomial,
  control=glmerControl(optimizer = 'optimx',
    optCtrl=list(method='bobyqa')))
```

In this chunk of code, the formula is made of the `Judgment`, which is the dependent variable (y_j in 2.1), the fixed effects, and the random variables. The relationship of the fixed effects is coded by astricts, which in this code indicates that the interactions among variables also need to be accounted for. In addition, the `binomial` indicates the categorical type of analysis.

Model output

Before heading towards the interpretation of the model's output, it would be crucial to have a better understanding of the model itself. As discussed in the previous section, a key part of the interpretation of the model in understanding the reference point. When illustrating the equation 2.1, it was mentioned that a is the intercept of the model or, to be more specific, is where all of the predictors are set to their 0 conditions. In our three-way model, the 0 conditions of the predictors are, in **Non-Verbal** element, which has two levels, *nonce* and *NV* (which stands for the meaningful elements), the *nonce* is set to the reference point. In terms of the **Light Verbs**, *?a:vardan*, 'to bring' is set to 0, and the rest of them have been compared according to that. And finally, the **Proto-Roles** were also coded as *Proto-Agent* is set to 0 and *Proto-Patient* is set to 1.

Following this explanation, in table 2.6, the *intercept* contains the condition where the Proto-Role is Proto-Agent, the Non-Verbal is Nonce, and the Light Verb is *?a:vardan*. The results of the model is based on the change of the *Odds Ratio* according to the baseline or reference point; namely, the other ratios are illustrated based on their difference with the reference point. table 2.6, also provides essential information about the random effects and how smoothly the model was run. Firstly, there were no warnings on convergence issues. Moreover, the variance of the random effects indicates that the random effects were chosen correctly since they are not exactly zero. The higher variability in the participants rather than those of the questions (items) indicates that the questions were slightly difficult for the participants, which resulted in more diverse answers. The other important column of the table considers the *Confidence interval*, in simple words, this factor indicates if there is any quasi or complete separation in the data. The results of our model indicate that *gereftan* 'to take', in our model is having a zero confidence interval which is a sign of a quasi-separation. Moving back to table 2.4, it can be seen that, this LV has not received any proto-patient entities. As a result, since it affects the whole regression model, to understand the model better, this LV has been omitted from the further steps of the analysis. An overall visualization of the data can be seen fig. 2.4.

Contrast Coding

The interpretation of the results in dummy coded analysis is not an easy and informative task. In order to have a better understanding of the results, scholars follow two main routes. One possible solution is using post-hoc analyses, such as estimated marginal means. Post-hoc analyses are in fact helpful in many cases but

Predictors	Judgment		
	Odds Ratio	Confidence interval	P
(Intercept)	7.69	4.29 – 13.78	<0.001
NV [NV]	3.67	1.25 – 10.82	0.018
LV [Dadan]	0.82	0.38 – 1.77	0.606
LV [daftar]	0.87	0.40 – 1.90	0.725
LV [Gereftan]	1.72	0.71 – 4.14	0.230
LV [Kardan]	1.17	0.52 – 2.65	0.707
LV [kefidan]	1.72	0.71 – 4.14	0.230
LV [fodan]	0.68	0.32 – 1.44	0.314
LV [Zadan]	1.92	0.78 – 4.75	0.156
protoroles [2]	0.07	0.04 – 0.14	<0.001
NV [NV] × LV [Dadan]	0.14	0.04 – 0.49	0.002
NV [NV] × LV [daftar]	0.04	0.01 – 0.14	<0.001
NV [NV] × LV [Gereft]	2546056.80	0.00 – 361118736388357919910423992280809472.00	0.667
NV [NV] × LV [Kardan]	0.59	0.14 – 2.56	0.482
NV [NV] × LV [kefidan]	0.98	0.17 – 5.50	0.979
NV [NV] × LV [fodan]	0.07	0.02 – 0.25	<0.001
NV [NV] × LV [Zadan]	0.06	0.02 – 0.25	<0.001
NV [NV] × protoroles [2]	0.09	0.03 – 0.30	<0.001
LV [Dadan] × protoroles [2]	1.90	0.77 – 4.70	0.163
LV [daftar] × protoroles[2]	1.44	0.58 – 3.57	0.436
LV [Gereft] × protoroles[2]	0.37	0.14 – 1.03	0.057
LV [Kardan] × protoroles[2]	0.86	0.33 – 2.21	0.750
LV [kefidan] × protoroles[2]	0.39	0.14 – 1.07	0.068
LV [fodan] × protoroles[2]	1.32	0.54 – 3.23	0.547
LV [Zadan] × protoroles[2]	0.28	0.10 – 0.79	0.016
(NV [NV] × LV [Dadan]) × protoroles [2]	13.84	3.14 – 60.91	0.001
(NV [NV] × LV [daftar]) × protoroles [2]	225.41	51.42 – 988.17	<0.001
(NV [NV] × LV [Gereft]) × protoroles [2]	0.00	0.00 – 13084975338310965657600.00	0.636
(NV [NV] × LV [Kardan]) × protoroles [2]	1.99	0.38 – 10.50	0.415
(NV [NV] × LV [kefidan]) × protoroles [2]	0.78	0.11 – 5.42	0.798
(NV [NV] × LV [fodan]) × protoroles [2]	111.41	25.76 – 481.88	<0.001
(NV [NV] × LV [Zadan]) × protoroles [2]	68.23	14.10 – 330.12	<0.001
Random Effects			
σ^2	3.29		
$\tau_{0question}$	0.03		
τ_{0ID}	0.36		
ICC	0.11		
N_{ID}	62		
$N_{question}$	64		
Observations	3968		
Marginal R ² / Conditional R ²	0.781 / 0.804		

Table 2.6: Output of the dummy-coded model

they come with the cost of having multiple analyses for a complete understanding of the data (Brehm and Alday, 2020). The other solution is using contrast coding, instead of dummy coding. In the previous section, it was mentioned that in dummy-coded analysis the intercept is the reference point and the comparisons are according to that point. In a contrast coded analysis, for example, the two-level variables instead of being coded, as (0,1) are coded in (-1,1). This recoding changes the intercept to the grand mean of the data. Namely, it indicates the average of all of the conditions according to the fixed and random effects, and the results are compared to this point. As mentioned in the previous section, our data frame is made of two two-level variables, and one eight-level variable, one of its levels was identified as quasi separation in the data and omitted; therefore, seven levels

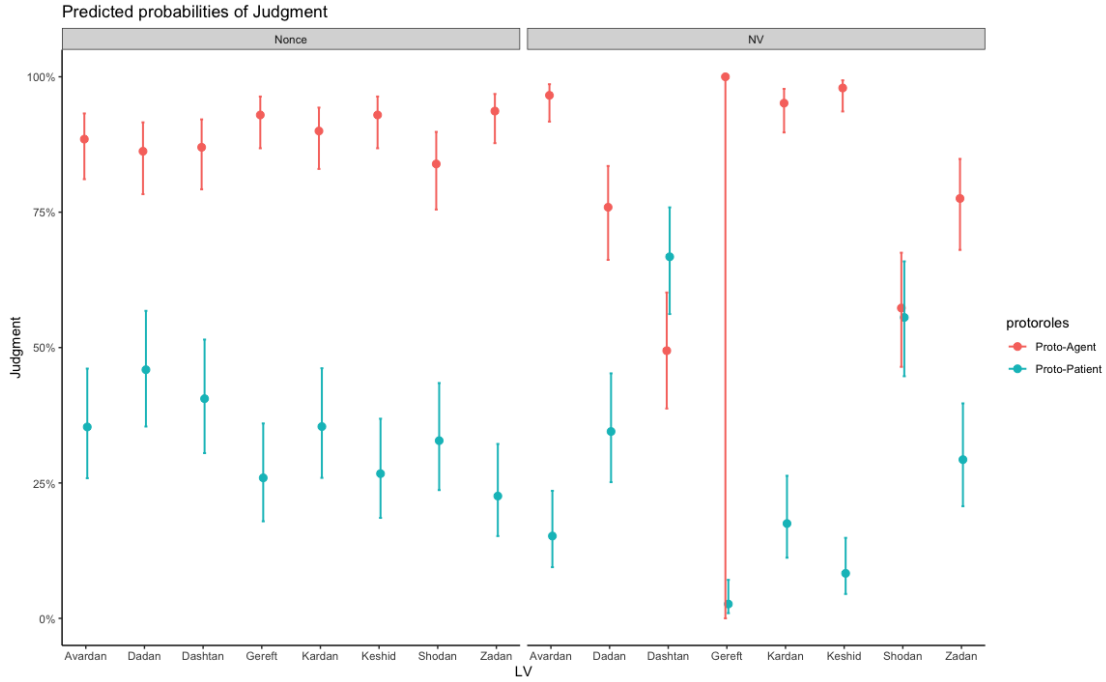


Figure 2.4: Predication of the proto-roles based on the light verbs

remained.

Following the tutorial of [Schad et al. \(2020\)](#), since we are interested in observing differences among the individual LVs, they were contrast-coded as sum or deviation contrast. In this contrast, as shown in eq. (2.2) each of the levels is compared with the grand mean of the data. Therefore, the first contrast compares the first condition to the average of the rest other levels, and this continues in all of the levels. The only level which is not present in any of the contrasts and is considered as the reference is set to be *zadan*, 'to hit'. Since in this research all of the levels of the LVs need to be compared and studied. The results of the contrast-coded analysis are not sufficient. Therefore, we will conduct a post-hoc analysis of the LVs to have a better understanding of the results.

$$\begin{pmatrix}
 & [, 1] & [, 2] & [, 3] & [, 4] & [, 5] & [, 6] \\
 Avardan & 1 & 0 & 0 & 0 & 0 & 0 \\
 Dadan & 0 & 1 & 0 & 0 & 0 & 0 \\
 Dashtan & 0 & 0 & 1 & 0 & 0 & 0 \\
 Kardan & 0 & 0 & 0 & 1 & 0 & 0 \\
 Keshid & 0 & 0 & 0 & 0 & 1 & 0 \\
 Shodan & 0 & 0 & 0 & 0 & 0 & 1 \\
 Zadan & -1 & -1 & -1 & -1 & -1 & -1
 \end{pmatrix} \quad (2.2)$$

The output of the contrast-coded model shown in table 2.7. In order to under-

stand the significance of the changes, *p value* has been reported, a value of $p < 0.05$, indicating a significant effect. Following this, the data, indicate that the main effect of the non-verbal and the proto-roles are significant. In addition, the interaction of the LVs and the proto-roles are significant in all of the cases. Furthermore, the interaction of the NV and the LVs on the proto-roles is significant, except for the LV2-nd contrast, with *p value* = 0.732.

Predictors	Judgment		
	<i>Odds Ratio</i>	<i>Confidence interval</i>	<i>P</i>
(Intercept)	1.81	1.51 – 2.18	< 0.001
NV	0.88	0.79 – 0.98	0.026
protoroles	0.25	0.23 – 0.28	< 0.001
NV×protoroles	1.02	0.92 – 1.13	0.732
LV1×protoroles	0.58	0.44 – 0.77	< 0.001
LV2×protoroles	1.54	1.26 – 1.89	< 0.001
LV3×protoroles	2.68	2.19 – 3.29	< 0.001
LV4×protoroles	0.64	0.49 – 0.83	0.001
LV5×protoroles	0.34	0.24 – 0.48	< 0.001
LV6×protoroles	2.16	1.77 – 2.63	< 0.001
(NV×LV1)×protoroles	0.54	0.41 – 0.71	< 0.001
(NV×LV2)×protoroles	1.04	0.85 – 1.27	0.732
(NV×LV3)×protoroles	2.07	1.69 – 2.54	< 0.001
(NV×LV4)×protoroles	0.64	0.49 – 0.83	0.001
(NV×LV5)×protoroles	0.51	0.36 – 0.71	< 0.001
(NV× LV6)×protoroles	1.74	1.43 – 2.12	< 0.001
Random Effects			
σ^2	3.29		
$\tau_{00question}$	0.04		
τ_{00ID}	0.32		
ICC	0.10		
N_{ID}	62		
$N_{question}$	56		
Observations	3472		
Marginal R^2 / Conditional R^2	0.428 / 0.485		

Table 2.7: Output of the contrast-coded model

Post-hoc Analysis

According to the outcome of the model, it can be assumed that there is an effective interaction between the LVs, NVs, and the proto-roles. In order to have the results confirmed as mentioned above, we perform a post-hoc analysis to investigate

the significance of the LVs and highlight the points of difference. In the post-hoc analysis, we observe the marginal means of the levels of the variables. This can highlight that, if the interaction of the NVs and the LVs with the proto-roles is not significant, in which condition this has happened, and what its motives could be. In order to perform the post-hoc analysis, we use the `emmeans` package, which calculates the marginal means, and returns the comparisons according to the desired conditions.

In the following chapter, I will discuss how the results of the model, and investigate how the different conditions have affected the judgments of the participants. In addition, I will discuss how the obtained results, are interpreted and can find an answer to our question according to our three hypotheses.

Chapter 3

Results

Thematic roles, as labels, and more specifically, agent and patient roles as schemas playing a critical role in language processing, have been studied in this research on the Persian complex predicate constructions. The case of complex predicates is interesting because they are made of two elements, and in theory, they have been said to play different roles in defining the thematic roles. It was explained that, in this research, the proto-role framework of [Dowty \(1991\)](#), has been used to investigate the contribution of these elements in allocating the roles to the subjects of the sentences. To tackle this question, a survey was run, and the results were analyzed in a linear mixed-effect model, enabling us to observe how different elements of the constructions contribute to this process.

In the literature on thematic roles, there are three main hypotheses considering the thematic roles of the CPrs. Firstly, there are compositional approaches such as the constructionist approach of ([Folli et al., 2005](#), among others) and the projectionist/lexicalist approach of ([Karimi-Doostan, 2005](#), among others). The former argue that the thematic role of the constructions is highly affected by the light verbs of the constructions, and in latter favors an effect of non-verbal element. In addition, the posterior-compositional approaches, such as those of ([Samvelian and Faghiri, 2014](#), among others), argue for a cumulative and interactive effect of the elements, arguing that a single type of light verb can only be meaningfully analyzed based on the context it appears in, and various non-verbal elements will co-occur with a light verb based on the semantic classes the light verb can fall into. Having all of these hypotheses in mind, we investigate the agent and patient roles in eight LV constructions in this research. This variety of the LVs can help us observe if there is a difference between them in allocating the roles. In order to observe the role of the NVs, we used nonce and meaningful NVs, expecting to observe changes in the patterns if they are effective in the readings. The linear mixed-effect model

also accounts for the interaction of the two elements, assisting us in finding out if these elements have an interactive role in the readings of their thematic roles and how their effectiveness is distributed.

In this chapter, the results of the model, have been interpreted and explained how they account for the hypotheses. In addition, we argue that the LVs do not seem to be interpreted in the same way, and they have different readings from each other.

3.1 The role of the Non-Verbal Element

Having observed the results of the survey in table 2.4, it can be seen that there is a change among the responses, in cases where the non-verbal element was meaningful or nonce. However, the degree of the change and the effect of the NVs on the proto-roles does not seem to be significant. As the results in table 2.5, shows the model with the two-way interaction of the NV and the proto-roles is less significant than the other models. This is also supported by the results of the contrast-coded model in table 2.7, reporting $\beta = 0.01774$, $p = 0.731603$. This can be seen in fig. 3.1, which shows that the slope of the change in judgments of proto-roles has not changed significantly, between the two conditions of the NVs. It can be concluded that the average of the judgments on the proto-roles is not directly affected by the NVs. This finding of the current research poses a challenge towards the lexicalist approach that accounts for the NVs to be the elements that form the event structure and denote the thematic role regardless of the LVs.

3.2 The role of the Light Verb Element

The role the LVs play in the judgments of the participants seems to be significantly high. As the results of the questionnaires in table 2.4, indicate that there is a difference among the results LVs have received. This can be seen in the two-way interactions of the LVs and the proto-roles reported by the model in table 2.7. To have a better vision of the effect of the LVs on the proto-roles, fig. 3.2 visualizes the differences between the average of the judgments among the LVs. It can be noticed that the slope between the LVs changes significantly. This observation confirms the hypothesis that the LVs affect the thematic roles of the complex predicates; however, it is still unclear if this is only because of the LVs, or it is because of the effect of both NVs and the LVs.

The comparison of the two-way model of the proto-roles and the LVs and the

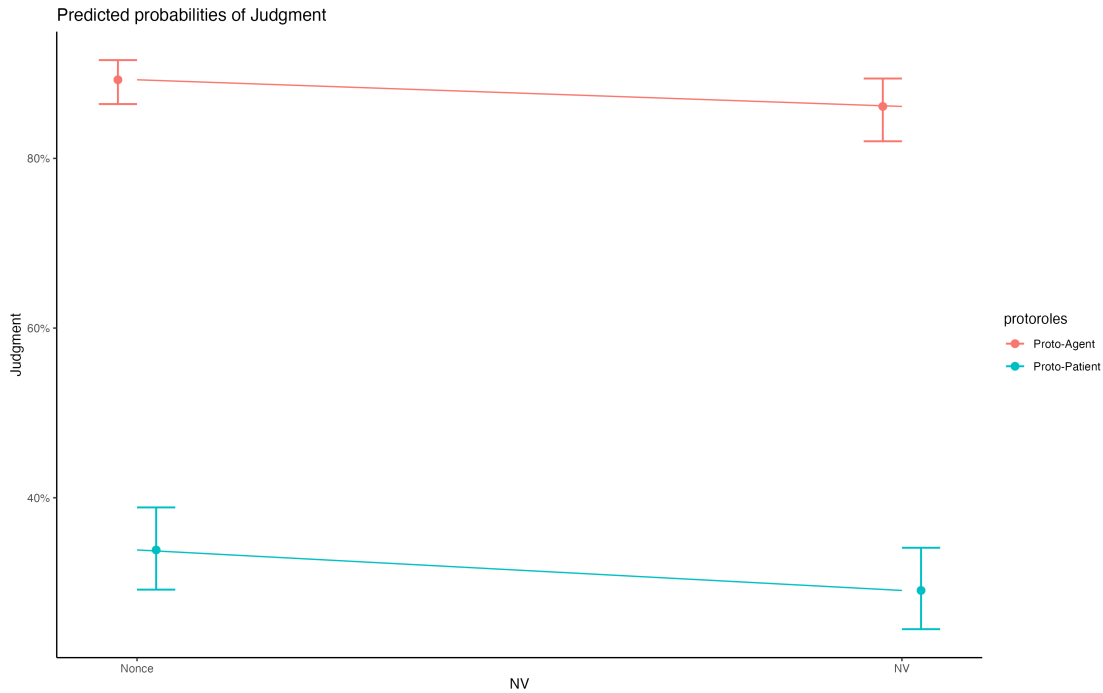


Figure 3.1: The interaction of the non-verbal and proto-roles

three-way model in table 2.5 suggests that the interaction of the LV and NVs on the judgments about the proto-roles is more significant, $p = < 0.001$, and the presence of the NVs critically affects the results. This confirms that perhaps the best way to account for the allocation of the thematic roles in the CPRs is to investigate them; and considering the allocation of the thematic role as result of the interactions of both of the elements.

3.3 The interaction of the Elements

An interactive role of the complex predicates in allocating the thematic roles seems to be the optimal hypothesis among the other hypotheses. Considering the results of the model in table 2.7, it can be seen that, all of the instances do not seem to have a similar degree of effectiveness. Therefore, it is needed to conduct a post-hoc analysis of the results to have a more comprehensive answer. In this research as mentioned in section 2.2.2 `emmeans` package is used for analyzing the expected marginal means for different conditions of the results.

First, let us observe the changes in the judgments of the proto-roles, regarding the effect of the NVs, for each of the LVs. This comparison enables us to notice the effect of the NVs, under different conditions. Having observed the fig. 3.3, it can be seen that, the slopes of the changes are not similar among the LVs, suggesting the

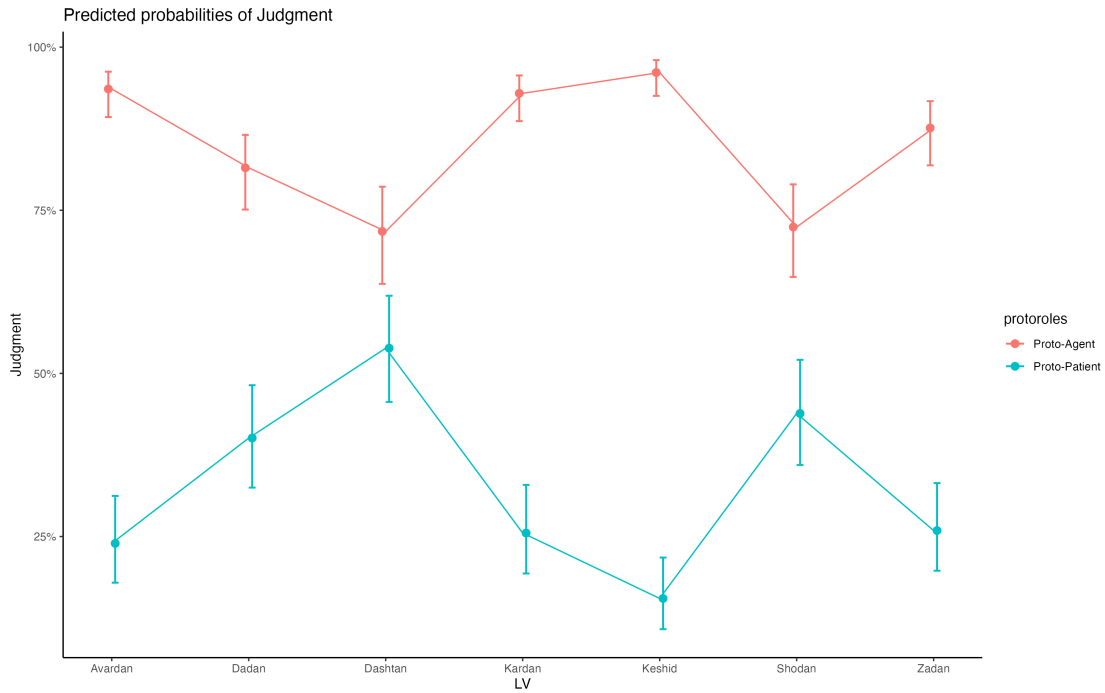


Figure 3.2: The interaction of the light verbs on judgments

NVs have different degrees of effectiveness in different LVs. As table 3.1 indicates, NV do not seem to affect *dadan*, 'to give' in neither agent nor patient roles. In addition, the effectiveness of the NVs in the decision of participants on the proto-agent role of *kardan*, to do, and *keshidan*, to pull, and the proto-patient role *zadan*, to hit, is not significant; although, in the rest of the conditions, NVs have reliably affected the judgments of the participants.

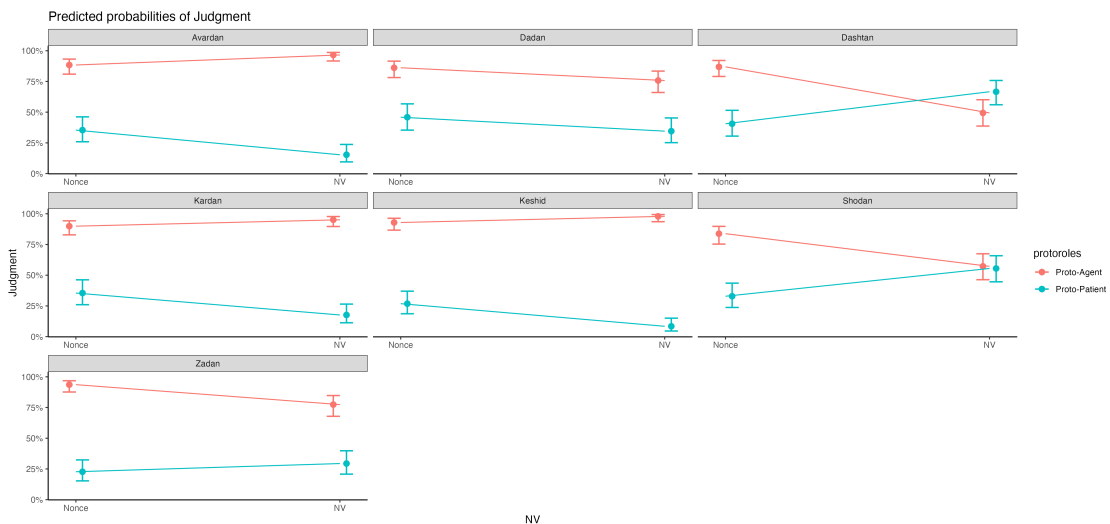


Figure 3.3: The interaction of the non-verbal and proto-roles on the judgements

Considering these results, let us observe how the allocation of the proto-roles

Light Verbs	Proto-roles	Results	
		Estimate	P value
ʔa:vardan 'to bring'	Proto-Agent	-1.302	0.0180
	Proto-Patient	1.107	0.0012
dadan 'to give'	Proto-Agent	0.687	0.0541
	Proto-Patient	0.474	0.1183
daftan, 'to have'	Proto-Agent	1.915	<0.001
	Proto-Patient	-1.073	0.0004
kardan, 'to do'	Proto-Agent	-0.777	0.1239
	Proto-Patient	0.943	0.0049
kefidan 'to pull'	Proto-Agent	-1.278	0.0635
	Proto-Patient	1.383	0.0005
ʃodan 'to become'	Proto-Agent	1.351	<0.001
	Proto-Patient	-0.933	0.0022
zadan 'to hit'	Proto-Agent	1.452	0.0008
	Proto-Patient	-0.349	0.2860

Table 3.1: Post-hoc Analysis on the effect of NVs

changes among the LVs, and what role we can expect, in different CPrs. As the graph in table 3.2, indicates, the difference between the thematic roles and the LVs regarding their NV is not similar. As it can be seen, in LVs of *ʔa:vardan*, 'to bring', *dadan* 'to give', *kardan*, 'to do', *kefidan*, 'to pull', and *zadan*, 'to hit', the estimates are all positive and the p values are smaller than 0.05; however, in *in daftan*, 'to have', there is a negative estimate $\beta = -0.7163$; suggesting that, the role of the subject can change in meaningful context and become more patient like. Considering the information from table 3.1, it can be said that the proto-role in this LV is reliably dependent on the NV, while in the absence of an NV, similar to the other LVs, it is more probable to have proto-agent as the subject. In addition, in *ʃodan*, 'to become', in meaningful instances, the difference between the proto-roles is not significant $p = 0.7906$; giving the impression that, in this LV the probability of having a proto-agent or proto-patient in subject is almost equal. This can be because of the meaning of the LV and the fuzzy difference of the proto-roles or the structure of the sentences, which were not considered in this research.

To sum up, based on the intuition obtained from the results of the three-way analysis, it can be argued that based on the findings of the survey, in the absence of a meaningful non-verbal element, there is a tendency in CPr constructions to have proto-agent properties in their subject position, this is also in harmony with the perspective of Dowty (1991), and the results of Kako (2006), and those of Reisinger et al. (2015). Although, In most of the LVs, this trend remains consistent

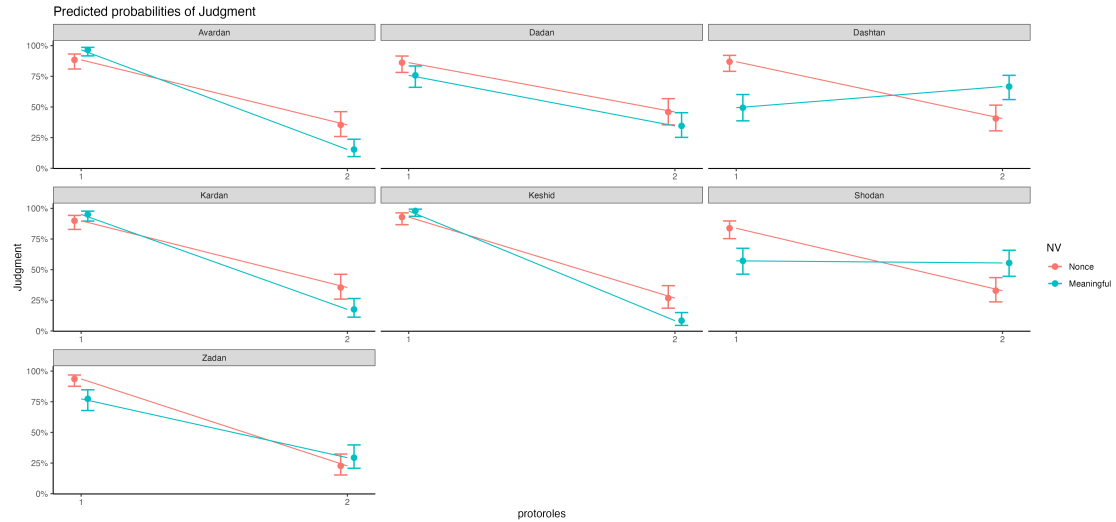


Figure 3.4: The interaction of the NV and proto-roles on the judgements of proto-roles

with the presence of the meaningful NV; in two of the constructions, to be specific in constructions with LV of *daftan*, 'to have', and *fodan*, 'to become', a change in the trend was observed; e.g., in the former, the role shifts from agent to the patient; and latter the proto-roles it takes become indistinguishable. The other important information data provides considers the fact that there are no light verbs that have a trend of patientivity in their subject when the effect of the NVs are absent. Furthermore, the differences in the degree of agentivity among LVs suggest that LVs are not semantically bleached and they affect the formation of the event structure; and are being processed by the speakers.

The results of the current study based on the observed trends favor an interactive, unequal role of the CPrs counterparts in allocating the thematic roles. In this regard, the results of this research support some of the arguments of the constructionalist approach and those of construction-based approach; while it cannot account for those of lexicalist approaches.

In the following chapter, I will conclude the results and explain how this research can answer the questions we asked at the beginning of this study, and how it affects the theoretical backgrounds on CPrs. In addition, I will explain the limitations of the research and, the possible areas for future studies and niches for further investigations.

Light Verbs	Results		
	Non-Verbal	Estimate	P value
ʔa:vardan 'to bring'	Nonce	2.6314	<.0001
	Meaningful	5.0410	<.0001
dadan 'to give'	Nonce	1.9918	<.0001
	Meaningful	1.7783	<.0001
daftan, 'to have'	Nonce	2.2715	<0.001
	Meaningful	-0.7163	0.0080
kardan, 'to do'	Nonce	2.7846	<.0001
	Meaningful	4.5052	<.0001
kefidan 'to pull'	Nonce	3.5726	<.0001
	Meaningful	6.2329	<.0001
ʃodan 'to become'	Nonce	2.3554	<0.001
	Meaningful	0.0705	0.7906
zadan 'to hit'	Nonce	3.9074	<.0001
	Meaningful	2.1065	<.0001

Table 3.2: Post-hoc Analysis on the effect of NVs and LV on proto-role judgements

Chapter 4

Conclusion and Discussion

This master thesis aimed to answer the question of how the thematic roles of the subjects of the Persian complex predicates (CPr) are allocated. The critical difference between these constructions with other constructions is that they are made of two components. Therefore, different theoretical approaches account for different characteristics of their components.

I started this research by introducing the notion of argument structure, defining it, and indicating how it is used in our everyday language understanding. Next, I reviewed the theories related to the CPrs. Firstly the lexical approaches toward the CPrs were reviewed. It was explained that from the perspective of scholars such as [Karimi-Doostan \(1997, 2005\)](#), the importance of the meaning is the key notion in the analysis. Hence, based on the definition of the LVs in [Grimshaw and Mester \(1988\)](#), e.g., they are semantically bleached elements, it has been relied on the characteristics of the non-verbal element in forming the argument structure and allocating the thematic roles. In this sense, the light verb element is only accounted for the syntactic processes of the CPrs. On the other hand, constructionist approaches such as [Folli et al. \(2005\)](#), and [Megerdooonian \(2001, 2012\)](#), shift the focus from NVs to the LVs, arguing that the role in construction can change based on the LVs they take. In this regard, even passivization is considered a change in the LVs. The last approach we reviewed in this research is the construction-based or usage-based approach, arguing that the CPrs, although have syntactically independent features, in the lexicon are represented as a single unit. In this regard, scholars such as [Goldberg \(1996\)](#); [Goldberg et al. \(2003\)](#), and [Samvelian and Faghiri \(2014\)](#) argue that an investigation of the CPrs needs to be conducted after their incorporation and *a posteriori*.

I continued with the review of the thematic roles, including two main perspectives of thematic role lists and generalized perspectives of them. Thereupon the cognitive

science accounts of the thematic roles were presented, and it was argued that the agent and patient roles indicate a universal bias being used in our everyday language understanding. Since the proto-role framework of Dowty (1991) mainly focuses on the clusters of agent and patient roles and provides a mechanism for investigating them, we moved to this framework and used it for analyzing the CPrs. Using this framework, we moved forward with the other language modeling and psycholinguistic studies.

To answer the main question of the research, we devised a questionnaire using eight light verbs. For observing how the effect of the NVs is, nonce NVs were used along with the meaningful instances. Following the models of thematic roles, we focused on the subject position of the constructions, asking the participants to fill in the blanks with two entities, derived from White et al. (2016), representing the proto-agent and proto-patient roles.

Taking the literature of the CPrs into consideration, we had three possible answers. Therefore, we tested three linear mixed effect models and drew a comparison among them. The results of the comparisons and the models suggested that the optimal model is three-way containing the thematic roles, NV, and the LV. This suggests that perhaps the construction-based approach can better model the thematic role allocation of the CPrs.

What do the results of this research tell us about the CPrs? I argue that, perhaps, the constituting elements of the CPrs, have an interaction with each other. This interaction is not equal and fixed among the components. As a result, I argue that the thematic role in CPrs is defined and shaped as a consequence of the interplay of these elements. This suggests that LVs are not semantically bleached; while they even affect Persian speakers' understanding of thematic roles, although their effectiveness is not similar among all and consistent. As a result of the model, a prediction of the thematic roles based on the effectiveness of the NV and LVs was proposed. In most of the cases, the difference between the proto-roles was significant in both of the NV conditions, indicating that the grammatical position and the LV seem to have a reliable effect on the allocation of the thematic roles, and tend to denote proto-agent characteristics; however, in cases such as *daftan*, 'to have', and *fodan*, 'to become' the results were not consistent and as discussed in the results chapter, in *daftan* the tendency changes to proto-patient when meaningful NV is present and in *fodan*, 'to have' the possible changes to almost 50%, which shows that the difference between proto-agent and proto-patient is not significant.

Considering the contributions of the results of the current research to theoretical studies, it can be argued that the results of our research pose against the lexicalist account of the CPrs since there was no direct effect of them on the judgments

regarding the proto-roles. While the perspectives of [Folli et al. \(2005\)](#), and [Samvelian and Faghiri \(2013a\)](#) seem to both, some revisions fit the results of the current study.

In terms of the constructionalist approach, it can be argued that, although the effect of the LVs in allocating the thematic roles of the CPrs is evident, and in the absence of an NV, they demonstrate a tendency to take the agentive role in their subjects; the constructionalist approach does not take the characteristics of NVs into account for the allocation of the thematic roles and cannot account for the effects of the NVs in LVs such as *daftan*, to have and *fodan*, to become. Considering the construction-based approach of [Samvelian and Faghiri \(2013a\)](#) and [Goldberg \(1996\)](#) although their account for the interaction of the NV and the LV seems to fit the results of this study better, this study did not find the evidence of having a significant role of the NV in all of the instances. In contrast, based on their approach, it was expected to have a significant effect on the NVs and LVs on the thematic roles in most of the instances.

This research is a step forward in studies related to the CPrs, and opens the space for new curiosities and future studies; however, it was not without limitations. Perhaps the questionnaire used in this research is not the most sophisticated gathering of the examples, and it was limited to a small number of NV and LV elements. Furthermore, it was the first time that it was being used in a research and its deficiencies in the sampling were unclear. One of the other limitations which is practically related to the notion of the syntax and semantic interface studies is that the Agentivity/Patientivity of a fully interpreted phrase is not necessarily the same as the Agentivity/Patientivity inherent in the lexical semantics of the predicate; therefore, there can be a degree of error in the features of the selected entities, and in the contexts, they were used. In addition, although the number of participants was close to the results of the power test, it could be higher and provide us with more judgments.

Due to the exploratory nature of this study, I hope that it will be replicated, and the results could be extended in future studies. For example, there are still relatively large numbers of LVs such as *xordan*, 'to collide', which were not included in the research. From the cognitive point of view, it is interesting to observe if the way children acquire the thematic roles of the CPrs is in harmony with the results of this experiment or if they take different accounts for learning the CPr elements and their correspondence to the thematic roles.

Appendix A

NVs and LVs used in the questionnaire

NV Element	NV Type / translation	Co-occurring LV
emtehan	exam	
edame	continue	
hatalmast	nonce	
bumak	nonce	
samar	outcome	<i>dadān</i> , 'to give'
zaḡr	suffer	
nast	nonce	
ḡegaf	nonce	

NV Element	NV Type / translation	Co-occurring LV
pust	peel	
bahane	excuse	
hatalmast	nonce	
ḡegaf	nonce	
emtehan	exam	<i>gereftān</i> , 'to take'
jad	memory	
ḡafe	nonce	
bumak	nonce	

NV Element	NV Type / translation	Co-occurring LV
gerje	cry	
bahane	excuse	
zoot	nonce	
hatalmast	nonce	<i>kardan</i> , 'to do'
ʃena	swim	
avaz	change	
si ʃar	nonce	
faber	nonce	

NV Element	NV Type / translation	Co-occurring LV
zaḡr	suffer	
ḡa ru	sweep	
ʃegaf	nonce	
faber	nonce	<i>kefidan</i> , 'to pull'
dard	pain	
da:d	shout	
afak	nonce	
ʃafe	nonce	

NV Element	NV Type / translation	Co-occurring LV
hefz	preserve	
hal	solve	
ʃegaf	nonce	
siʃar	nonce	<i>ʃodan</i> , 'to become'
moteva ḡeh	understand	
avaz	change	
siʃar	nonce	
hatalmast	nonce	

NV Element	NV Type / translation	Co-occurring LV
da:d	shout	
hatalmast	nonce	
zoot	nonce	
ahang	song	
ɔ̃garu	sweep	<i>zadan</i> , 'to hit'
pang	nonce	
ma:zid	nonce	
kapak	mold	

NV Element	NV Type / translation	Co-occurring LV
bahane	excuse	
gir	catch	
hatalmast	nonce	
ʃegaf	nonce	
taʃrif	come	<i>avardan</i> , 'to bring'
be jad	memroy	
mand	nonce	
ʃafe	nonce	

NV Element	NV Type / translation	Co-occurring LV
zoot	excuse	
ʃeto	catch	
negah	nonce	
samar	nonce	
bumak	come	<i>daftan</i> , 'to have'
afak	memroy	
sora	nonce	
ʃafe	nonce	

Table A.1: distribution of non-verbals and the light verbs in the questionnaire

Appendix B

List of the Questions and the Filler entities

- *emtehan dadan*, 'to take an exam'
 - proto-agent entity: *irani-ha*, 'Iranians'
 - proto-patient entity: *in taqirrat*, 'these changes'
- *tfizi ra edame dadan*, 'to continue something'
 - proto-agent entity: *jeki az hamijan-e mantfester junajted*, 'one of the Manchester united supporters'
 - proto-patient entity: *qaza-je batfe*, 'kid meal'
- *hatalmast dadan*, 'to nonce have'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *tfizi ra bumak dadan*, 'to nonce have something'
 - proto-agent entity: *mohaqeq-an*, 'researchers'
 - proto-patient entity: *daneftgah-ha*, 'universities'
- *tfizi ra pust gereftan*, 'to peel something'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *bahane gereftan*, 'to take an excuse'

- proto-agent entity: *kimija*, 'Kimia (proper noun)'
- proto-patient entity: *je kar-e fegeft-angiz*, 'a marvelous job'
- *hatalmast gereftan*, 'to nonce take'
 - proto-agent entity: *mohaqeq-an*, 'researchers'
 - proto-patient entity: *danesgah-ha*, 'universities'
- *tfizi ro tfegaf gereftan*, 'to nonce take something'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *qaza-je batfe*, 'kid meal'
- *gerye kardan*, 'to cry'
 - proto-agent entity: *mohaqeq-an*, 'researchers'
 - proto-patient entity: *in taqirrat*, 'these changes'
- *tfizi ro bahane kardan*, 'to make an excuse of something'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *zoot kardan*, 'to nonce do'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *tfizi ro hatalmast kardan*, 'to nonce do something'
 - proto-agent entity: *mo?alem-in*, 'teachers'
 - proto-patient entity: *in taqirrat*, 'these changes'
- *zacgr kefidan*, 'to suffer'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *qaii ra qarar kefidan*, 'to sweep somewhere'
 - proto-agent entity: *mo?alem-in*, 'teachers'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'

- *tfegaf kefidan*, 'to nonce pull'
 - proto-agent entity: *masʔud va reza*, 'Masoud and Reza (proper noun)'
 - proto-patient entity: *in taqirrat*, these changes
- *tfizi ro faber kefidan*, 'to nonce pull something'
 - proto-agent entity: *jeki az hamijan-e mantfester junajted*, 'one of the Manchester united supporters'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *tfizi ro hefz fiodan*, 'to memorize something'
 - proto-agent entity: *comonist-ha*, 'communists'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *hal fiodan*, 'to become solved'
 - proto-agent entity: *comonist-ha*, 'communists'
 - proto-patient entity: *in moskelat*, 'these problems'
- *tfegaf fiodan*, 'to nonce become'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *tfizi ro sifar fiodan*, 'to nonce become something'
 - proto-agent entity: *mohaqq-an*, 'researchers'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *hatalmast zadan*, 'to nonce hit'
 - proto-agent entity: *comonist-ha*, 'communists'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *kasi ro zoot zadan*. 'to nonce hit someone'
 - proto-agent entity: *jeki az hamijan-e mantfester junajted*, 'one of the Manchester united supporters'
 - proto-patient entity: *in taqirrat*, these changes
- *ahang zadan*, 'to sing'

- proto-agent entity: *masʔud va reza*, 'Masoud and Reza (proper noun)'
- proto-patient entity: *in taqiirat*, these changes
- *ɕaii ro ɕaru zadan*, 'to sweep somewhere'
- proto-agent entity: *masʔud va reza*, 'Masoud and Reza (proper noun)'
- proto-patient entity: *in taqiirat*, these changes
- *bahane avardan*, 'to bring an excuse'
- proto-agent entity: *mohaqq-an*, 'researchers'
- proto-patient entity: *daneſgah-ha*, 'universities'
- *kasi ro gir avardan*, 'to find someone'
- proto-agent entity: *jeki az hamijan-e manfester junajted*, 'one of the Manchester united supporters'
- proto-patient entity: *qaza-je batfe*, 'kid meal'
- *ɕizi ro hatalmast avardan*, 'to nonce bring something'
- proto-agent entity: *kimija*, 'Kimia (proper noun)'
- proto-patient entity: *qaza-je batfe*, 'kid meal'
- *ɕegaſ avardan*, 'to nonce bring'
- proto-agent entity: *jeki az hamijan-e manfester junajted*, 'one of the Manchester united supporters'
- proto-patient entity: *je kar-e ſegeft-angiz*, 'a marvelous job'
- *ɕizi ro zoot daſtan*, 'to nonce have something'
- proto-agent entity: *jeki az hamijan-e manfester junajted*, 'one of the Manchester united supporters'
- proto-patient entity: *qaza-je batfe*, 'kid meal'
- *ſeto daſtan*, 'to nonce have'
- proto-agent entity: *foruſande*, 'salesperson'
- proto-patient entity: *vaziſe-je ma*, 'our duty'
- *ɕizi ro negah daſtan*, 'to keep something'

- proto-agent entity: *kimija*, 'Kimia (proper noun)'
- proto-patient entity: *vazife-je ma*, 'our duty'
- *samar daftan* 'to be fruitful'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *in taqirrat*, 'these changes'
- *samar dadan*, 'to become fruitful'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *kasi ro zaqr dadan*, 'to suffer someone'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *nest dadan*, 'to nonce give'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *tfizi ro chegafdadan*, 'to nonce give something'
 - proto-agent entity: *mo?alem-in*, 'teachers'
 - proto-patient entity: *in taqirrat*, 'these changes'
- *emtehan gereftan*, 'to give an exam'
 - proto-agent entity: *mohaqeq-an*, 'researchers'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *tfizi ro jad gereftan*, 'to learn something'
 - proto-agent entity: *irani-ha*, 'Iranians'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *chafe gereftan*, 'to nonce take'
 - proto-agent entity: *comonist-ha*, 'communists'
 - proto-patient entity: *in taqirrat*, 'these changes'

- *tfizi ro bumak gereftan*, 'to nonce take something'
 - proto-agent entity: *jeki az hamijan-e mantfester junajted*, 'one of the Manchester united supporters'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *fena kardan*, 'to swim'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *bavar-ha-moon*, 'our beliefs'
- *tfizi ro avaz kardan*, 'to change something'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *sifar kardan*, 'to nonce do'
 - proto-agent entity: *jeki az hamijan-e mantfester junajted*, 'one of the Manchester united supporters'
 - proto-patient entity: *qaza-je batfe*, 'kid meal'
- *kasi ro faber kardan*, 'to nonce do someone'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *dard kefidan*, 'to suffer from pain'
 - proto-agent entity: *mohaqeq-an*, 'researchers'
 - proto-patient entity: *danešgah-ha*, 'universities'
- *tfizi ro dad kefidan*, to shout something
 - proto-agent entity: *mohaqeq-an*, 'researchers'
 - proto-patient entity: *in taqirrat*, 'these changes'
- *afak kefidan*, 'to nonce pull'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *tfizi ro chafe kefidan*, 'to nonce pull something'

- proto-agent entity: *masʔud va reza*, 'Masoud and Reza (proper noun)'
- proto-patient entity: *daneʔgah-ha*, 'universities'
- *ʔfizi ro motevajah sodan*, 'to understand something'
 - proto-agent entity: *irani-ha*, 'Iranians'
 - proto-patient entity: *in moʔkelat*, 'these problems'
- *avaz sodan*, 'to change'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *sifar sodan*, 'to nonce become'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *ʔfizi ro hatalmat sodan*, 'to nonce become something'
 - proto-agent entity: *foruʔande*, 'salesperson'
 - proto-patient entity: *qaza-je batʔe*, 'kid meal'
- *pang zadan*, 'to nonce hit'
 - proto-agent entity: *foruʔande*, 'salesperson'
 - proto-patient entity: *vazife-shun*, 'their duty'
- *ʔfizi ro maazid zadan*, 'to nonce hit something'
 - proto-agent entity: *foruʔande*, 'salesperson'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *kapak zadan*, 'to mold'
 - proto-agent entity: *foruʔande*, 'salesperson'
 - proto-patient entity: *qaza-je batʔe*, 'kid meal'
- *ʔfizi ro dad zadan*, 'to shout something'
 - proto-agent entity: *mohaqq-an*, 'researchers'
 - proto-patient entity: *daneʔgah-ha*, 'universities'

- *taʃrif avardan*, 'to come'
 - proto-agent entity: *forufande*, 'salesperson'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'
- *ʔizi ra be jad avardan*, 'to remember something'
 - proto-agent entity: *comonist-ha*, 'communists'
 - proto-patient entity: *vazife-je ma*, 'our duty'
- *ʔizi ro mand avardan*, 'to nonce bring'
 - proto-agent entity: *irani-ha*, 'Iranians'
 - proto-patient entity: *daneʃgah-ha*, 'universities'
- *chafe avardan*, 'to nonce bring'
 - proto-agent entity: *masʔud va reza*, 'Masoud and Reza (proper noun)'
 - proto-patient entity: *in taqirrat*, 'these changes'
- *bumak daftan*, 'to nonce have'
 - proto-agent entity: *moʔalem-in*, 'teachers'
 - proto-patient entity: *daneʃgah-ha*, 'universities'
- *ʔizi ro afak daftan*, 'to nonce have something'
 - proto-agent entity: *mohaqeq-an*, 'researchers'
 - proto-patient entity: *daneʃgah-ha*, 'universities'
- *kasi ro soraG daftan*, 'to know someone'
 - proto-agent entity: *irani-ha*, 'Iranians'
 - proto-patient entity: *daneʃgah-ha*, 'universities'
- *edame daftan*, 'to have continue'
 - proto-agent entity: *kimija*, 'Kimia (proper noun)'
 - proto-patient entity: *je kar-e segeft-angiz*, 'a marvelous job'

Bibliography

- Alishahi, A. and Stevenson, S. (2010). A computational model of learning semantic roles from child-directed language. *Language and Cognitive Processes*, 25(1):50–93.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):390–412.
- Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.
- Baker, M. C. (1988). *Incorporation: A Theory of Grammatical Function Changing*. University of Chicago Press.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Brehm, L. and Alday, P. M. (2020). A decade of mixed models: It’s past time to set your contrasts. In *the 26th architectures and mechanisms for language processing conference (amlap 2020)*.
- Chang, N. (2004). Putting meaning into grammar learning. In *Proceedings of the Workshop on Psycho-Computational Models of Human Language Acquisition*, pages 19–26.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT press.
- Chomsky, N. (1981). Lectures on government and binding, foris, dordrecht. *Chomsky Lectures on Government and Binding 1981*.
- Croft, W. (1986). Surface subject choice of mental verbs. In *annual meeting of the Linguistic Society of America, New York*.

- Dabir-Moghaddam, M. (1997). Compound verbs in persian.
- Dowty, D. (1989). On the semantic content of the notion of ‘thematic role’. In *Properties, types and meaning*, pages 69–129. Springer.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *language*, 67(3):547–619.
- Fillmore, C. J. (1968). The case for case, dins. In Bach, E. and Harms, R., editors, *Universals in Linguistic Theory*. Holt, Rinehart, and Winston.
- Folli, R., Harley, H., and Karimi, S. (2005). Determinants of event type in persian complex predicates. *Lingua*, 115(10):1365–1401.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Goldberg, A. E. (1996). Words by default: Optimizing constraints and the persian complex predicate. In *Annual Meeting of the Berkeley Linguistics Society*, volume 22, pages 132–146.
- Goldberg, A. E. et al. (2003). Words by default: The persian complex predicate construction. *Mismatch: Form-function incongruity and the architecture of grammar*, 1:17–146.
- Grimm, S. (2006). The lattice of case and agentivity.
- Grimshaw, J. and Mester, A. (1988). Light verbs and θ -marking. *Linguistic inquiry*, pages 205–232.
- Gruber, J. S. (1965). *Studies in lexical relations*. PhD thesis, Massachusetts Institute of Technology.
- Hafri, A., Papafragou, A., and Trueswell, J. C. (2013). Getting the gist of events: recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, 142(3):880.
- Hale, K. and Keyser, S. J. (2002). *Prolegomenon to a theory of argument structure*. MIT press.
- Hale, K., Keyser, S. J., et al. (1993). On argument structure and the lexical expression of syntactic relations. *The view from Building*, 20(20):53–109.
- Jackendoff, R. (1972). Semantic interpretation in generative grammar.

- Jackendoff, R. (1983). *Semantics and cognition*, volume 8. MIT press.
- Jackendoff, R. (1987). The status of thematic relations in linguistic theory. *Linguistic inquiry*, 18(3):369–411.
- Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4):434–446. Special Issue: Emerging Data Analysis.
- Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, 101(1):1–42.
- Karimi, S. (1997). Persian complex verbs: Idiomatic or compositional. *LEXICOLOGY-BERLIN-*, 3:273–318.
- Karimi-Doostan, G. (1997). *Light verb constructions in Persian*. PhD thesis, Ph. D. dissertation, University of Essex.
- Karimi-Doostan, G. (2005). Light verbs and structural case. *Lingua*, 115(12):1737–1756.
- Karimi-dustan, G. and Eshaghi, M. (2019). Systematic light verbalization in persian language. *Researches in Linguistics*, 11(2):157–186.
- Lebani, G., Bondielli, A., and Lenci, A. (2015). You are what you do. an empirical characterization of the semantic content of the thematic roles for a group of italian verbs.
- Lebani, G. E. and Lenci, A. (2018). A distributional model of verb-specific semantic roles inferences. *Language, cognition, and computational models*, 118.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.
- Levin, B. (2019). On dowty’s ‘thematic proto-roles and argument selection’.
- Levin, B. and Rappaport Hovav, M. (2005). *Argument Realization*. Research Surveys in Linguistics. Cambridge University Press.
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198):226–228.
- McRae, K., Ferretti, T. R., and Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, 12(2-3):137–176.

- Megerdooonian, K. (2001). Event structure and complex predicates in persian. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 46(1-2):97–125.
- Megerdooonian, K. (2012). The status of the nominal in persian complex predicates. *Natural Language & Linguistic Theory*, 30(1):179–216.
- Nilipour, R. (2000). Agrammatic language: Two cases from persian. *Aphasiology*, 14(12):1205–1242.
- Nilipour, R. and Raghibdoust, S. (2001). Manifestations of aphasia in persian. *Journal of Neurolinguistics*, 14(2):209–230.
- Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms. *Language*, 70(3):491–538.
- Pollet, S. (2012). *Grammaire des prédicats complexes: les constructions nom-verbe (Collection langues et syntaxe)*. Lavoisier.
- Pustejovsky, J. and Batiukova, O. (2019). *The Lexicon*. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Reisinger, D., Rudinger, R., Ferraro, F., Harman, C., Rawlins, K., and Van Durme, B. (2015). Semantic proto-roles. *Transactions of the Association for Computational Linguistics*, 3:475–488.
- Rissman, L. and Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic bulletin & review*, 26(6):1850–1869.
- Rosch, E. and Lloyd, B. B. (1978). Cognition and categorization.
- Sadeghi, A. A. (1993). On denominative verbs in persian. *Farsi Language and the Language of Science*, pages 236–246.
- Samvelian, P. and Faghiri, P. (2013a). Introducing PersPred, a syntactic and semantic database for Persian complex predicates. In *Proceedings of the 9th Workshop on Multiword Expressions*, pages 11–20, Atlanta, Georgia, USA. Association for Computational Linguistics.
- Samvelian, P. and Faghiri, P. (2013b). Re-thinking compositionality in persian complex predicates. In *Annual Meeting of the Berkeley Linguistics Society*, volume 39, pages 212–226.
- Samvelian, P. and Faghiri, P. (2014). Persian complex predicates: How compositional are they? *Semantics-Syntax Interface*, 1(1):43–74.

- Schad, D. J., Vasishth, S., Hohenstein, S., and Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110:104038.
- Seraji, M., Ginter, F., and Nivre, J. (2016). Universal Dependencies for Persian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2361–2365, Portorož, Slovenia. European Language Resources Association (ELRA).
- Shabani-Jadidi, P. (2014). *Processing compound verbs in Persian: A psycholinguistic approach to complex predicates*. Leiden University Press.
- Strickland, B. (2017). Language reflects “core” cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive Science*, 41(1):70–101.
- Van Valin, R. D. (2001). Semantic macroroles in role and reference grammar.
- Van Valin, R. D. and LaPolla, R. J. (1997). *Syntax: Structure, Meaning, and Function*. Cambridge University Press, Cambridge.
- Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, 66(2):143–160.
- White, A. S., Reisinger, D., Sakaguchi, K., Vieira, T., Zhang, S., Rudinger, R., Rawlins, K., and Van Durme, B. (2016). Universal decompositional semantics on universal dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723.
- White, A. S., Stengel-Eskin, E., Vashishtha, S., Govindarajan, V., Reisinger, D. A., Vieira, T., Sakaguchi, K., Zhang, S., Ferraro, F., Rudinger, R., et al. (2019). The universal decompositional semantics dataset and decomp toolkit. *arXiv preprint arXiv:1909.13851*.
- Wickham, H. and Girlich, M. (2022). *tidyr: Tidy Messy Data*. <https://tidyr.tidyverse.org>, <https://github.com/tidyverse/tidyr>.