



# How much is determined by Syntax? An empirical Approach to the Position of the Direct Object in Persian

Pegah Faghiri, Pollet Samvelian

## ► To cite this version:

Pegah Faghiri, Pollet Samvelian. How much is determined by Syntax? An empirical Approach to the Position of the Direct Object in Persian. 9th Iranian Conference on Linguistics, Allameh Tabataba'i University, Department of Linguistics Feb 2015, Tehran, Iran. pp. 1419-1435. halshs-01220381v1

**HAL Id: halshs-01220381**

**<https://shs.hal.science/halshs-01220381v1>**

Submitted on 26 Oct 2015 (v1), last revised 23 Dec 2015 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## How Much is Determined by the Syntax? An Empirical Approach to the Position of the Direct Object in Persian

Pegah Faghiri & Pollet Samvelian  
*Université Sorbonne Nouvelle – Paris 3*

### Abstract

Several theoretical studies (e.g. Browning & Karimi 1994; Karimi 2003; Ghomeshi 1997; Ganjavi 2007) claim that *rā*-marked (definite and/or specific) direct objects (DOs) occupy a higher position than their non-*rā*-marked (indefinite non-specific) counterparts. One of the main arguments to support this claim is the unmarked relative order between the direct and the indirect object (IO), which is broadly assumed to be IO–DO<sub>[–rā]</sub>–V and DO<sub>[+rā]</sub>–IO–V (e.g. Mahootian 1997; Rasekhmahand 2004; Roberts 2009). In this paper, we provide empirical evidence against such a dichotomous view.

We have conducted a corpus-based study (Faghiri & Samvelian, 2014) and experimental follow up studies (Faghiri *et al.*, 2014; Faghiri, forthcoming) to investigate ordering preferences between the DO and the IO in the preverbal domain. In addition to the realization of the DO, we have taken into account other potentially influential factors such as relative length, givenness, collocationality and lexical bias, via mixed-effect regression modeling, in line with key empirical studies on word order variations (e.g. Wasow 2002).

The data reveal that while *rā*-marked DOs show a strong preference to appear before the IO, among different non-*rā*-marked DOs, *i.e.* bare nouns *ketāb*, bare noun with modifiers *ketāb-e akkāsi* and indefinite/numeral NPs *ye ketāb(-i)*, only bare nouns show a strong preference for adjacency to the verb. Interestingly, indefinite (non-*rā*-marked) DOs show a clear preference for the inverse, grouping thus with *rā*-marked DOs. Moreover, extra syntactic factors such as relative length also play a significant role on these ordering preferences. Accordingly, we argue that the ordering preferences observed for different types of DOs is best reflected by a continuum based on the degree of conceptual and/or discourse accessibility. Consequently, any structural account of the latter would lead to wrong predictions.

We furthermore examine other arguments provided to support a dual-position account and present some counterexamples that question their validity. Thus, in line with Samvelian (2001), we refute a dual syntactic position account of the DO.

### 1. Introduction

The position of the direct object (DO) in Persian has received a lot of attention, not only with respect to Persian grammar, but also from a theoretical point of view, especially on “scrambling” (cf. Karimi, 2005). Differential object marking (DOM) is responsible for this especial treatment. In Persian, a definite and/or specific DO is always marked. Unmarked DOs vary from NPs carrying an overt (indefinite) determination to bare

nouns, that is, nouns appearing in their singular bare form carrying no determination, quantification or modification. On this spectrum, the properties of marked DOs seem to contrast sharply with those of bare nouns, especially with respect to the word order. This has brought many scholars, mainly in the generative framework, to suggest that marked and unmarked DOs are structurally different and do not occupy the same syntactic position. Furthermore, Persian, due to its limited number of simplex verbs, is characterized by the significant presence of complex predicates (CPr), the combination of a non-verbal element (mostly a bare noun) with a verb<sup>1</sup>. Thus, the tight semantic relation between the bare noun and the verb in a CPr has reinforced the idea that the comparable strong semantic bond that exists between the verb and its unmarked DO, contrary to its marked DOs, may receive a structural explanation (cf. Karimi, 1999, 2003).

In this paper, building on a thorough empirical investigation based on corpus and experimental data, we argue against a dual syntactic position for DOs in Persian. A crucial argument put forward in previous studies to support the dual view is the neutral word order in ditransitive constructions. In a neutral word order, marked DOs precede the IO while unmarked DOs are adjacent to the verb. The data we report here show that this claim is only partially true and that the unmarked word order in these constructions cannot receive a structural explanation. Indeed, an accessibility-based account can explain ordering preferences between the two objects crosslinguistically and beyond syntactic constraints.

We also review other syntactic contrasts put forward to justify a dual-position hypothesis, and provide counterexamples to express our disagreement with the corresponding grammatical judgments. We conclude, in line with Samvelian (2001), that a dual-position view of DOs is flawed.

The remainder of this paper is organized as follows. In the next section, we present the dual syntactic position hypothesis, focusing on the account provided by Karimi (2003). In section 2, we argue against this hypothesis and in section 3, we provide the results of our empirical studies. In section 5, we discuss the role of accessibility on the ordering preferences between the DO and the IO.

## 2. Dual Syntactic Position Accounts of the DO

Several theoretical studies, mainly in the generative framework, have suggested a dual-position account of the DO in Persian depending on its *rā*-markedness (Browning & Karimi 1994; Ghomeshi, 1997; Karimi, 2003; Ganjavi, 2007, among others). Despite significant differences, these studies claim that *rā*-marked DOs do not occupy the same syntactic position as their non-*rā*-marked counterparts. Whether this difference is base-generated or is a result of an “object shift” or scrambling, it is generally assumed that in spell out a *rā*-marked DO appears in a higher position than a non-*rā*-marked DO. To

---

<sup>1</sup> Many scholars assume a categorical distinction between the combination of a verb and a noun in a CPr and ordinary verb-complement combinations, Samvelian (2001, 2012) is a notable exception.

illustrate this view, we briefly discuss Karimi (2003) as a representative of the dual syntactic position account.

Karimi (2003) suggests that specific and nonspecific objects have different structural properties and therefore do not occupy the same syntactic position. A specific DO occupies the position of the specifier of VP, while a non-*rā*-marked DO occupies a lower position, that is, the position sister to the verb (under the V'). The two positions suggested by Karimi (2003, p.105)<sup>2</sup> are given in (1). She additionally assumes that a specific DO, definite or indefinite, is always followed by *-rā*, while its nonspecific counterpart lacks *=rā*.

- (1) a. [VP DP<sub>[+Specific]</sub> [V' PP V]]  
 b. [VP [V' PP [V' DP<sub>[-Specific]</sub> V]]]

Karimi (2003) mentions different instances of syntactic asymmetries<sup>3</sup> between the two types of DOs to support her claim, considering that the unmarked (relative) word order between the DO and the IO is crucial among others. The following examples from Karimi (2003, pp.91–92) illustrate this asymmetry<sup>4</sup>.

- (2) a. Kimia aqlab barā mā še'r mi-xun-e  
 Kimea often for us poem IPFV<sup>5</sup>-read-3SG  
 'It is often the case that Kimea reads poetry for us.'
- b. Kimia aqlab barā mā ye še'r az Hafez mi-xun-e  
 Kimea often for us a poem of Hafez IPFV-read-3SG  
 'It is often the case that Kimea reads a poem by Hafez for us.'
- (3) a. Kimia aqlab hame=ye še'r-hā=ye tāze-aš=ro barā mā mi-xun-e  
 Kimea often all=EZ<sup>6</sup> poem-PL-EZ fresh=3SG=DOM for us IPFV-read-3SG  
 'It is often the case that Kimea reads all her new poems for us.'
- b. Kimia aqlab ye še'r az Hafez=ro barā mā mi-xun-e  
 Kimea often a poem of Hafez=DOM for us IPFV-read-3SG  
 'It is often the case that Kimea reads a (particular) poem by Hafez for us.'

Karimi, furthermore, assumes that a nonspecific or, in other words, a non-*rā*-marked DO can be separated from the verb, that is, can undergo scrambling, only if it has a contrastive focus, as illustrated by examples in (4).

<sup>2</sup> Note that in a more recent work, Karimi (2005) proposes a revised version of her Two Object Position Hypothesis (TOPH), in which both objects are base-generated in the same position, that is, the position under the v'. The specific object shifts into the specifier of vP position in order to receive its interpretation.

<sup>3</sup> Semantic asymmetries between the two DOs are discussed as well. However, this issue is beyond the scope of our study.

<sup>4</sup> These examples are modified according to the glossing and transcription used in this paper.

<sup>5</sup> Glosses follow the Leipzig Glossing Rules ([www.eva.mpg.de/lingua/resources/glossing-rules.php](http://www.eva.mpg.de/lingua/resources/glossing-rules.php)). The following non-standard abbreviations are used for clarity: DOM = differential object marking; EZ = Ezafe.

<sup>6</sup> The Ezafe, realized as an enclitic, links the head noun to its modifiers and to the possessor NP (see Samvelian 2007).

- (4) Kimia aqlab (ye) ketāb=e dāstān barā bačče-hā mi-xun-e  
 Kimea often (a) book=EZ story for child-PL IPFV-read-3SG  
 ‘Kimea often reads (a) STORY-BOOK for children (rather than a poetry book).’

### 3. Against Dual Accounts

Despite their differences, all the studies that suggest a structural difference between the two types of DOs are built on the claim that in a neutral or unmarked word order, *rā*-marked DOs precede the IO, while their non-*rā*-marked counterparts follow the IO. In this paper, we discuss empirical evidence that undermines this claim, see section 4. The data show that: 1) The distribution of the unmarked relative order between the two objects is not dichotomous and is best reflected by a continuum on the basis of the degree of determination of the DO. 2) Extra syntactic factors like relative length play a significant role on the ordering preferences between the two objects. As Faghiri *et al.* (2014) argue, these two factors converge on the ground of conceptual and/or discourse accessibility (see section 5) and hence the relative order between the two objects can be accounted for on the basis of accessibility, which applies beyond hard-core syntactic constraints and is valid crosslinguistically. This implies that any structural account of ordering possibilities between the two objects would generate wrong predictions.

While the unmarked word order asymmetry is the backbone argument of the dual syntactic position hypothesis, other syntactic asymmetries are also put forward to support this view. Namely, Karimi (2003) mentions a difference of behavior between the two DOs with respect to licensing parasitic gaps and binding anaphors, as well as the impossibility to appearance together in a coordinate construction. The empirical data that we discuss in this paper focus only on the unmarked word order asymmetry. Nevertheless, we will briefly go through other mentioned asymmetries in this section. We believe that empirical studies, namely in experimental paradigms such as acceptability judgment ratings<sup>7</sup>, should be conducted to investigate the empirical validity of these arguments as well. Meanwhile, we express our disagreement with Karimi’s (2003, 1999) grammaticality judgments by providing counterexamples. Also, see Samvelian (2001) for additional arguments against a dual-position hypothesis.

- 1) According to Karimi (1999, p.704) only *rā*-marked (specific) DOs can license parasitic gaps, as illustrated by (5). Yet, the sentences in (6) and (7) sound perfect to us. The oddness of (5) may be due to the fact that the verb is in the past tense and the sentence denotes an specific accomplished event where it is expected for the DO to be known to the speaker and hence a bare DO is not felicitous.

- (5) a. Kimea [<sub>DP</sub> in ketāb=ro]<sub>i</sub> [<sub>CP</sub> qabl-az inke pro e<sub>i</sub> be-xun-e] be man dād  
 Kimea this book=DO before that SBJV-read-3SG to I gave  
 ‘Kimea game me this book before reading (it)’  
 b. \*Kimea [<sub>DP</sub> ketāb]<sub>i</sub> [<sub>CP</sub> qabl-az inke pro e<sub>i</sub> be-xun-e] be man dād

<sup>7</sup> Note that these queries are difficult to pin down in corpus data, as well as in sentence production paradigms.

- (6) man ketāb<sub>i</sub> qabl-az inke pro e<sub>i</sub> be-xun-am be kasi (kādo) ne-mi-d-am  
 I book before that SBJV-read-1SG to s.o. (gift) NEG-IPFV-give-1SG  
 ‘I wouldn’t offer someone a book/books (as a gift) before reading (it/them)’
- (7) man faqat [ye ketāb]<sub>i</sub> qabl-az inke pro e<sub>i</sub> be-xun-am be Sara dāde=am  
 I only a book before that SBJV-read-1SG to Sara given=1SG  
 ‘I have only given one book before reading (it) to Sara’
- (8) man lebās<sub>i</sub> bedun=e inke pro e<sub>i</sub> emtehān kon-am ne-mi-xar-am  
 I cloth without=EZ that try do-1SG NEG-IPFV-buy-1SG  
 ‘I wouldn’t buy a cloth without trying (it)’
- (9) man [ye lebās]<sub>i</sub> bedun=e inke pro e<sub>i</sub> emtehān kon-am xarid-am va  
 I a cloth without=EZ that try do-1SG bought-1SG and  
 hičvaqt ham na-pušid-am=eš  
 never too NEG-wore-1SG=3SG  
 ‘I only bought one cloth without trying (it) and I never wore it.’

Note that, even if we agree with Karimi’s grammatical judgments in (5), there is no need for positing a structural difference between DOs. Since, as Karimi herself mentions in a former article on parasitic gaps (1999, p.707), the ungrammaticality of such examples can be accounted for on the basis of semantic mismatch between the licensing chain and the parasitic gap chain, since pronouns, here the *pro* in the parasitic gap, are specific.

2) Karimi claims that non-*rā*-marked (nonspecific) DOs cannot bind an anaphor:

- (10) a. man se-tā bačče-hā=ro be hamdige mo'arrefi kard-am  
 I three-cl child-PL=DOM to each-there introduce did-1SG  
 ‘I introduced the three children to each other.’
- b. \*man se-tā bačče be hamdige mo'arrefi kard-am
- (11) a. Kimea<sub>i</sub> [ye dānešju]<sub>k</sub> be hamšāgerdi-hā=š<sub>i/\*k</sub> t<sub>k</sub> mo'arrefi kard  
 Kimea a student to classmate-PL=3SG introduce did  
 ‘Kimea introduced a student to her classmates.’
- b. Kimea<sub>i</sub> [ye dānešju=ro]<sub>k</sub> be hamšāgerdi-hā=š<sub>i/k</sub> mo'arrefi kard

Here again, we believe that the ungrammaticality of (10) is due to semantic mismatch (and not to syntactic constraints). The anaphor *hamdige* ‘each other’, is a definite pronoun, yet in (10) its antecedent is an indefinite nonspecific NP. The same observation holds in (11). A nonspecific NP is not a felicitous antecedent for a definite pronoun in the same clause. Karimi (2003, p.113) suggests that because nonspecific DOs are part of the predicate, they cannot enter binding relations. Yet, a nonspecific NP cannot bind a definite/specific pronoun in the same clause, whether we consider it as a part of the predicate or not.

3) For Karimi (2003, p.103), the two types of DOs cannot appear together in a coordinate construction, see (12) below. We disagree with Karimi's grammatical judgments. We consider example (13) grammatical. In fact, ordering preferences in coordinate constructions need a thorough empirical investigation. We believe that the issue raised by (12) is rather the choice of the order between the coordinates. It seems that in these constructions the *rā*-marked DO is preferred in the final position of the coordination.

(12) \*man diruz in aks=ro va ketāb xarid-am  
I yesterday this picture=DOM and book bought-1SG

(13) man diruz šam'=o ye rumizi=o in tāblo=ro xarid-am  
I yesterday candle=and a tablecloth=and this painting=DOM bought-1SG  
'Yesterday, I bought this painting a nap and some candles.'

On the basis of this evidence, in line with Samvelian (2001), we conclude that a dual syntactic position account of the DO is flawed.

The dual-position hypothesis draws a binary distinction between specific/definite DOs and nonspecific/indefinite DOs. The data, as we will discuss in section 4, show that a binary classification depending on *rā*-marking is not sufficient to address the issue at stake. Namely, non-*rā*-marked DOs contrary to what is suggested in the above-mentioned studies do not have a homogenous behavior. This is not surprising given the fact that these DOs do not constitute a homogenous class with respect to determination. Hence, it is necessary to take a closer look to the classification of Persian NPs with respect to the definiteness and/or the degree of determination. Moreover, we will briefly review the differential object marking (DOM) in Persian, which as highlighted by almost all scholars, cannot be accounted for on the basis of specificity and/or definiteness only.

### 3.1. Persian NPs: A Fine-grained Classification

In formal Persian, there is no overt marker for definiteness; only indefiniteness is marked. Furthermore, Persian has what Corbett (2000) calls a general number, expressed by the singular form. This means that in Persian the number is not specified for a bare singular noun<sup>8</sup>. These properties have some bearings on the readings of NPs. In the remainder of this section, we will discuss the following NP types: bare and bare-modified nouns, indefinite/quantified NPs, and definite NPs.

#### 3.1.1. Bare and Bare-modified Nouns

Bare nouns are non-specified for number and have a nonspecific reading, which can be generic (14) as well as existential (15).

(14) gorg yek heyvān=e vahši va darande ast  
wolf a animal=DOM wild and predator is

---

<sup>8</sup> Nouns occurring alone in their bare singular form may correspond either to a definite and/or anaphoric NP or to nouns without any determination or quantification. By "bare noun", we only refer to the latter.

‘The wolf is a wild and predator animal.’

- (15) diruz gorg be deh āmade bud  
yesterday wolf to village come was  
‘Yesterday, a wolf/wolves came to the village.’

Bare-modified nouns only differ from bare nouns by the presence of a (restrictive) modifier.

- (16) diruz sag=e hār be deh āmade bud  
yesterday dog=EZ mad to village come was  
‘Yesterday, a mad dog/mad dogs came to the village.’

### 3.1.2. Indefinite NPs

Indefinite NPs can have either a specific or a nonspecific existential reading. These NPs, contrary to bare nouns, are always specified for number. Indefiniteness is overtly marked in Persian, realized by the enclitic =*i*, as in (17), by the cardinal *ye(k)* ‘one’, as in (18b), or by the combination of these two determiners, as in (18c). Numerals or other indefinite quantifiers, as in (18), also form indefinite NPs.

- (17) a. gorg=*i* zuze mi-kešid  
wolf=INDF howl IPFV-pulled  
b. yek gorg zuze mi-kešid  
a wolf howl IPFV-pulled  
c. yek gorg=*i* zuze mi-kešid  
a wolf=INDF howl IPFV-pulled  
‘A (any/certain) wolf was howling.’
- (18) čand(=tā)/se(=tā) gorg zuze mi-kešid-and  
few(=CLF)/three(=CLF) wolf howl IPFV-pulled-3PL  
‘A few/three wolves were howling.’

### 3.1.3. Definite NPs

Definite NPs can either be formed by different definite determiners, like demonstratives, or by no overt determiner, as in (19). Furthermore, bare plural nouns generally trigger a definite reading, as in (20). Note, however, that the plural marking is not incompatible with the indefinite determination =*i* or *yek*, as in (21) (for a discussion of plural marking and definiteness, see Ghomeshi 2003).

- (19) (in) šišē emruz šekast  
(this) glass today broke  
‘This/The glass broke today.’
- (20) šišē-hā emruz šekast-and  
glass-PL today broke-3PL  
‘The (\*Some) glasses broke today.’
- (21) yek šišē-hā=*i* heyn=e asbābkeši šekast-and  
a glass-PL=INDF during=EZ move broke-3PL



‘Some (of the) glasses broke during the move.’

### 3.2. Differential Object Marking

Persian displays differential object marking (DOM), realized by the enclitic =*rā*. Definite and/or specific direct objects are necessarily *rā*-marked. Consequently, non-*rā*-marked direct objects receive an indefinite nonspecific reading, as illustrated by (22). DOM is not incompatible with the indefinite determination, as illustrated by (23). An indefinite NP like *ketāb=i* when *rā*-marked will receive a specific reading.

- (22) Maryam ketāb xarid vs. Maryam ketāb=*rā* xarid  
Maryam book bought Maryam book=DOM bought  
‘Maryam bought a book/some books.’ vs. ‘Maryam bought the book’

- (23) Maryam ketāb=*i* xarid vs. Maryam ketāb=*i*=*rā* xarid  
Maryam book=INDF bought Maryam book=INDF=DOM bought  
‘Maryam bought a book.’ vs. ‘Maryam bought a (specific) book’

- (24) ketāb=*rā* mi-xān-and  
book=DOM IPFV-read-3PL  
‘A book, one reads it.’ or ‘A book is meant to be read’

- (25) emruz=*rā* dars mi-xān-am  
today=DOM lesson IPFV-read-1SG  
‘Today, I will study’

Nevertheless, *rā*-marking cannot be accounted for on the basis of definiteness and specificity only (cf. e.g. Dabir-Moghaddam, 1992), as illustrated by (24). Furthermore, the use of the enclitic =*rā* is not limited to DOM. It is also used to mark discourse prominence for other non-subject functions, as in (25). Meanwhile, a more detailed discussion is beyond the scope of the present study (for further discussions see Lazard 1982; Dabir-Moghaddam 1992; Meunier & Samvelian 1997; Roberts *et al.* 2009, among others). As far as this study is concerned, we associate *rā*-marking with the highest degree of determination for an NP in the DO position, as well as with the most prominent or salient constituent (excluding the subject) with respect to the information structure of the sentence.

### 4. Data-driven Evidence Against Dual Accounts

In this section, we discuss empirical evidence that undermines a dual-position hypothesis DOs in Persian. First, we show that contrary to what is largely admitted in the literature, the unmarked relative word order between the DO and the IO does not solely depend on *rā*-marking. Second, we show that relative length in terms of number of words between the two objects has a significant effect on the word order. This observation clearly shows that the word order between the two objects cannot be accounted for purely syntactically.

#### 4.1. The Unmarked Word Order

It is generally assumed that in unmarked, canonical or neutral word order ditransitive constructions in Persian *rā*-marked DOs precede the IO while non-*rā*-marked DOs follow the IO (e.g. Browning & Karimi 1994; Ghomeshi, 1997; Mahootian 1997; Karimi, 2003; Rasekhmahand 2004; Ganjavi, 2007; Roberts *et al.* 2009). While this hypothesis joins the prescriptions of reference grammars on the canonical word order, it is interesting to note that a different view, which considers the indefinite marker *-i*, also exists. Givi Ahmadi & Anvari (1995, p.305), for instance, state that *rā*-marked DOs should precede the IO, *i*-marked (non *rā*-marked) indefinite DOs can either follow or precede the IO, and non-*rā*-marked (non *i*-marked) DOs should follow the IO.

Lambrecht (1996) highlights that unmarked or neutral word orders, contrary to marked word orders, are not specified for a particular discourse function and can be used in any information structure. Consequently, having more distributional freedom, a neutral word order is the word order that has a greater overall frequency of occurrence. The empirical research that we report here sticks to Lambrecht's (1996) observation and thus is built on the following assumption: the unmarked word order is the one that everything being equal is significantly more frequent than the others.

The distribution of different ordering possibilities between the two objects that is, DO-IO-V and IO-DO-V, in corpus<sup>9</sup> and the follow up experimental studies (Faghiri & Samvelian 2014; Faghiri *et al.* 2014; Faghiri, forthcoming) reveals that the unmarked word order for indefinite non-*rā*-marked DOs is DO-IO-V and not the inverse as it is largely admitted. In fact, the DO-IO-V word order was significantly more frequent in corpus as well as in experimental data, respectively at a rate of 77% and 68%, for these DOs.

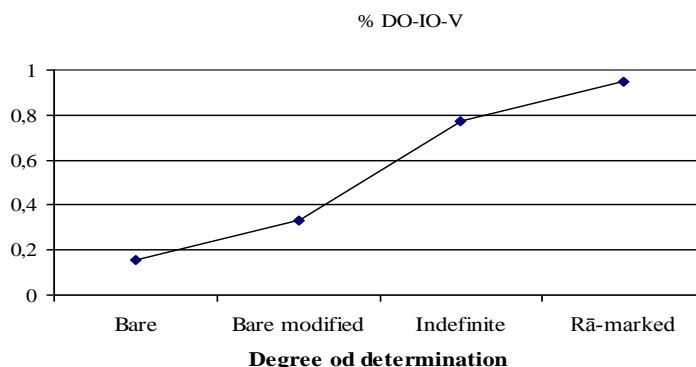
Furthermore, as illustrated by Figure 1 below<sup>10</sup>, the distribution of the word order with respect to different types of DOs, that is, *rā*-marked, indefinite, bare-modified and bare nouns<sup>11</sup>, reflects a continuum on the basis of the degree of determination of the DO. The more a DO is determined, the more it is likely to be separated from the verb, or, in other words, the less a DO is determined, the more it is likely to be adjacent to the verb.

---

<sup>9</sup> The dataset sampled out of the Bijankhan corpus (<http://ece.ut.ac.ir/dbrg/bijankhan/>) and contains 905 token, that is, verb-final ditransitive sentences. For more details, see Faghiri & Samvelian 2014.

<sup>10</sup> The percentages presented in this figure correspond to the corpus data (Faghiri & Samvelian 2014). Note that we also have the ratios for bare-modified and indefinite DOs in experimental data, but they do not differ from those represented here considerably.

<sup>11</sup> In a preliminary analysis, Faghiri & Samvelian (2014) observed that the marked vs. unmarked classification of DOs is inadequate, especially concerning non-*rā*-marked DOs. Accordingly, they defined a more fine-grained classification of the latter depending on the degree of determination, separating those carrying a determination, that is, indefinite DOs, from those carrying none, that is, bare and bare-modified DOs, see section 6.



**Figure 1 Word order and degree of determination of the DO**

#### 4.2. The Effect of Relative Length

Another crucial fact that the empirical data discussed here have revealed is the effect of the relative length (in terms of number of words) between the two objects on the word order. Indeed, corpus data as well as follow up experiments have found a statistically significant effect of the relative length corresponding to the “long-before-short” preference. For instance, in a sentence completion experiment with indefinite DOs (Faghiri *et al.*, 2014), the participants ordered the IO before the DO significantly more frequently when the IO was longer than the DO, and *vice versa*. The same effect is observed in a similar experiment with bare-modified DOs (Faghiri, forthcoming). The “long-before-short” preference is also observed in corpus data for these DOs, contrary to *rā*-marked and bare DOs, for which this factor does not play a significant role. In the corpus data, *rā*-marked almost exclusively precede the IO (at the rate of 95%), irrespective of the relative length. Note that, the relative length is not relevant for bare nouns, given that they are by definition shorter than the IO (for a detailed presentation of this study see Faghiri & Samvelian, 2014).

The “long-before-short” preference observed in the preverbal domain in Persian, lines up with observations on another verb-final language, Japanese for instance, for which the “long-before-short” preference is confirmed in corpus (e.g. Hawkins, 1994) as well as by experimental studies (e.g. Yamashita & Chang, 2001).

Note that this preference contradicts the largely admitted universality of the “short-before-long” preference observed mainly in Germanic languages like English and German, in accordance (and reinforcing) accessibility-based incremental models of sentence production. The latter claim that the linear order of constituents reflects the order in which they become available for production, as long as grammar rules do not intervene (e.g. Garrett, 1980; Bock & Levelt, 1994; Kempen & Harbusch, 2003). Constituents that become available at an earlier point in time can occupy an earlier linear position than constituents emerging later. This view is indeed fully compatible with the

“short-before-long” principle<sup>12</sup> (e.g. Stallings *et al.*, 1998; Arnold *et al.*, 2000; Wasow, 2002).

Yamashita & Chang (2001) provide experimental evidence for “long-before-short” preference in sentence production in Japanese<sup>13</sup> and proposed a production-oriented account of these conflicting ordering. The authors suggest that acknowledging language-specific differences in sentence production is the key to a uniform account of word order preferences. They assume that the sensitivity of sentence production system to conceptual vs. formal factors can be seen as being language-specific. The production system of Japanese, they argue, is more sensitive to conceptual factors than to form-related ones, contrary to English. This is because Japanese is a far less “rigid” language than English. Japanese has a fairly free word order and allows null pronouns. English, in contrast, has a fairly strict word order that requires all arguments to be overtly present (Yamashita & Chang, 2001, p.54). Moreover, in English Heavy-NP shift happens in the postverbal domain, where it is shown that the verb exerts strong influence, contrary to the preverbal domain (Stallings *et al.*, 1998). These syntactic constraints presumably increase the effect of form-related factors over more conceptual ones.

Long constituents have competing properties. They are semantically richer, due to the extra lexical material which makes them more salient and increases their overall accessibility in the conceptual arena. At the same time, in the form arena, the extra lexical content makes them slower to process and hence less accessible. The authors consequently conclude that the Japanese sentence production system, more sensitive to conceptual factors, favours placing long constituents before shorter ones, while in English, more sensitive to form-related factors, placing short constituents before longer ones is favoured.

Like Japanese, Persian is an SOV language with a fairly free word order and null pronouns. Hence, this language share all properties singled out by Yamashita & Chang (2001) to motivate opposite length-based shifts in Japanese and English. Consequently, assuming longer constituents are lexically richer and hence more salient and conceptually accessible, it is possible to attribute this “long-before-short” preference in the preverbal domain in Persian to the more important influence of conceptual factors, comparing to form-related ones (Faghiri & Samvelian, 2014; Faghiri *et al.*, 2014).

## 5. The Role of Accessibility

The empirical evidence presented here shows that the relative order between the DO and the IO in the preverbal domain in Persian depends on two factors: the degree of

---

<sup>12</sup> Short simple constituents can be processed and formulated faster and thus become available for production sooner than long and/or complex ones.

<sup>13</sup> The mirror-image preference in head-initial and head-final languages was first observed by Hawkins (1994, 2004), who proposed a dependency-based distance-minimizing principle in terms of a theory of parsing efficiency. He proposed the *Early Immediate Constituent (EIC)* principle to account for these seemingly contradictory preferences in head-final and head-initial languages. Note that since Persian is not a consistently head-final language like Japanese, EIC fails to provide adequate predictions for Persian, as illustrated by Faghiri & Samvelian (2014).

determination of the DO and its length. As argued in Faghiri *et al.* (2014, p.233), these two independent factors contribute both to the conceptual accessibility of the DO.

A closer look into the interaction between the relative length and the DO type shows that these two factors can be combined.

- Marked DOs, uncontroversially, strongly prefer to the DO-IO order, and bare DOs strongly prefer the IO-DO order, regardless of length.
- Indefinite DOs, present a moderate preference for the DO-IO order, which increases significantly for longer DOs.
- Bare-modified DOs can be viewed as longer counterparts of bare DOs. They are lexically richer and therefore, even though they display the same degree of discourse givenness as bare DOs, are more salient than the latter and hence conceptually more accessible. They show a less strong preference for the IO-DO order than bare DOs, which decreases for longer DOs.

Recall from section 3.2 that for an NP in the DO position in Persian, *rā*-markedness corresponds to the highest degree of determination, bare (and bare-modified for that matter) DOs to the lowest degree, and indefinite DOs to an intermediate degree. One can safely assume that this continuum, established on the basis of the degree of determination, reflects a hierarchy of discourse givenness and/or prominence (cf. Gundel *et al.* 1993).

Accordingly, bearing in mind the analysis of the “long-before-short” preference provided by Yamashita & Chang (2001), it is possible to establish a continuum based on the increasing degree of conceptual accessibility, combining discourse givenness/prominence and lexical salience, from the strong preference of bare DOs for the IO-DO-V order to the strong preference of *rā*-marked DOs for the DO-IO-V order.

## 6. Conclusion

In this paper, we have argued against the widespread dual syntactic position hypothesis for DOs in Persian, which assumes that DOs do not occupy the same syntactic position depending on their markedness. This assumption has been generally built on the claim that the neutral word order, marked DOs precede the IO, while unmarked DO follow it. Our corpus and experimental studies, adopting the methodology in recent key studies in word order variations, showed that this claim is flawed. Namely, we have shown that the distribution of the unmarked word order with respect to the DO type does not correspond to a dual behavior depending on *rā*-marking. It rather reflects a continuum based on the conceptual and discourse accessibility of the DO.

We have additionally questioned other arguments put forward in support of the syntactic asymmetry of specific vs. nonspecific DOs, namely licensing parasitic gaps or binding anaphors and the impossibility for the two DOs to appear in coordinate constructions.

Hence, not only, there is no empirical ground for positing two distinct syntactic positions for DOs in Persian, but also such a hypothesis makes wrong predictions on word order presences.

## References

- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3). 435–483.
- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language* 76(1). 28–55.
- Bock, J. Kathryn. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89(1). 1–47.
- Bock, J. Kathryn & Willem Levelt. 1994. Language production: Grammatical encoding. In Morton A. Gernsbacher (ed.), *Handbook of psycholinguistics*, 945–984. New York: Academic Press.
- Bossong, Georg. 1985. *Empirische Universalienforschung : differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen : Gunter Narr Verlag.
- Browning, Marguerite. & Ezat Karimi. 1994. Scrambling to object position in Persian. In N. Cover, H. van Riemsdijk (eds.), *Studies on scrambling*, 61–100.
- Corbett, Greville G. 2000. *Number*. Cambridge University Press.
- Dabir-Moghaddam, Mohammad. 1992. On the (in) dependence of syntax and pragmatics: Evidence from the postposition -ra in Persian. In *Cooperating with written texts: The pragmatics and comprehension of written texts*, 549–573. Mouton de Gruyter.
- Faghiri, Pegah. Forthcoming.
- Faghiri, Pegah & Pollet Samvelian. 2014. Constituent ordering in Persian and the weight factor. In Christopher Pinon (ed.), *Empirical issues in syntax and semantics 10*, 215–232.
- Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2014. Accessibility and Word Order: The Case of Ditransitive Constructions in Persian. In Stefan Müller (ed.): *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar*, University at Buffalo, 217–237. Stanford, CA: CSLI Publications.
- Ganjavi, Shadi. 2007. *Direct objects in persian*. University of Southern California.
- Garrett, Merrill F. 1980. Levels of processing in sentence production. *Language production* 1. 177–220.
- Ghomeshi, Jila. 1997. Topics in persian vps. *Lingua* 102(2). 133–167.
- Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia linguistica* 57(2). 47–74.
- Givi Ahmadi, Hassan & Hassan Anvari. 1995. *Dastur zabāne fārsi [Persian grammar]*. Mo’assese farhangi Fātemi.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

- Hawkins, John A. 1994. A performance theory of order and constituency. Cambridge University Press.
- Hawkins, John A. 2004. Efficiency and complexity in grammars. Oxford University Press.
- Karimi, Simin. 1999. A Note on Parasitic Gaps and Specificity. *Linguistic Inquiry* 30. 704-713.
- Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Blackwell Publishing.
- Karimi, Simin. 2005. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Blackwell Publishing.
- Kempen, Gerard & Karin Harbusch. 2003. Word order scrambling as a consequence of incremental sentence production. In Holden Härtl & Heike Tappe (eds.), *Mediating between concepts and grammar*, 141–64. Mouton De Gruyter.
- Lambrecht, Knud. 1996. Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Cambridge University Press.
- Lazard, Gilbert. 1982. Le morphème *rā* en persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris* 77(1). 177–208.
- Mahootian, Shahrzad. 1997. *Persian*. New York: Routledge.
- Meunier, Annie & Pollet Samvelian. 1997. La postposition *rā* en persan: son rôle dans la détermination et sa fonction discursive. *Cahiers de Grammaire* 25. 187–232.
- Rasekhmahand, Mohammad. 2004. *Jāygāh maf'ul mostaqim dar fārsi* [The position of the direct object in Persian]. *nāme farhangestān* 6:56–66.
- Roberts, John R, Behrooz Barjasteh Delforooz & Carina Jahani. 2009. A study of Persian discourse structure. *Acta Universitatis Upsaliensis*.
- Samvelian, Pollet. 2001. Le statut syntaxique des objets nus en persan. *Bulletin de la Société de Linguistique de Paris* 96(1). 349–388.
- Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian *ezafe*. *Journal of Linguistics* 43(3). 605–645.
- Samvelian, Pollet. 2012. *Grammaire des prédicats complexes : les constructions nom-verbe*. Hermès–Lavoisier.
- Stallings, Lynne M., Padraig G. O'seaghdha & Maryellen C. MacDonald. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language* 39(3). 392 – 417.
- Wasow, Thomas. 2002. Postverbal behavior. *CSLI*.
- Yamashita, Hiroko & Franklin Chang. 2001. 'Long before short' preferences in the production of a head final language, *Cognition* 81(2): B45-B55.