

How much structure is needed: The case of the Persian VP

Pegah Faghiri

Université Sorbonne Nouvelle Paris 3

Pollet Samvelian

Université Sorbonne Nouvelle Paris 3

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure
Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan
Müller (Editors)

2016

CSLI Publications

pages 236–254

<http://csli-publications.stanford.edu/HeadLex/2016>

Keywords: Persian, phrase structure, Differential Object Marking (DOM), empirical/usage-based methods

Faghiri, Pegah, & Samvelian, Pollet. 2016. How much structure is needed: The case of the Persian VP. In Arnold, Doug, Butt, Miriam, Crysmann, Berthold, King, Tracy Holloway, & Müller, Stefan (Eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 236–254. Stanford, CA: CSLI Publications.



Abstract

The aim of this paper is to tease apart two available views of the VP in Persian. The prevailing view of the Persian VP initially suggested in generative studies assumes a hierarchical structure with two object positions, mainly motivated by the existence of differential object marking in Persian. Building on quantitative studies, we revisit this hierarchical view and show that it is not born out by the data. A flat structure view of the VP, on the contrary, is in line with the data.

1 Introduction

In this paper we address the issue of the syntactic structure of the VP in Persian, an SOV language with mixed head direction (e.g. head-initial in NP, PP and CP), free word order in the clausal domain¹ and null pronouns.²

Previous generative studies on Persian VP have suggested a hierarchical structure that is motivated by the existence of differential object marking (DOM), which requires marking of definite direct objects (DOs) by the enclitic =*rā* (Karimi, 1990; Browning & Karimi, 1994; Ghomeshi, 1997; Karimi, 2005; Ganjavi, 2007). These studies have claimed that *rā*-marked and unmarked DOs display several syntactic and semantic asymmetries, for which this hierarchical view provides a straightforward account. Following insights from studies such as Diesing (1992), unmarked DOs have been assumed to be VP internal while *rā*-marked DOs are VP external. The higher syntactic position of the latter explains word order preferences in di-

[†]We would like to thank the audience at the HeadLex16 Conference (Polish Academy of Sciences, Warsaw, 24-29 July, 2016) for their questions, suggestions and insightful comments. We would also like to thank the two anonymous reviewers for their helpful comments on the earlier version of this paper. This work is supported by a public grant funded by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (reference: ANR-10-LABX-0083).

¹While SOV is the canonical word order in Persian, all other possibilities can also occur:

- a. Puyān Sepide=*rā* did
Puyan Sepideh=DOM see.PST.3SG
‘Puyan saw Sepideh.’
- b. Sepide=*rā* Puyān did (OSV)
- c. Puyān did Sepide=*rā* (SVO)
- d. Sepide=*rā* did Puyān (OVS)
- e. did Puyān Sepide=*rā* (VSO)
- f. did Sepide=*rā* Puyān (VOS)

²The following pair of examples illustrates the possibility to have covert arguments in subject and object positions in Persian, contrary to languages like English.

- | | | | |
|----|--|----|--|
| a. | Puyān Sepide= <i>rā</i> did?
Puyan Sepide=DOM see.PST.3SG
‘Did Puyan see Sepideh?’ | b. | na na-did
No NEG-see.PST.3SG
‘No, he did’t see her.’ |
|----|--|----|--|

transitive constructions. It also accounts for other asymmetries concerning binding and scope.

Recently, experimental and corpus-based studies have established that the most essential argument on which this view is built, namely ordering preferences, does not hold (Faghiri & Samvelian, 2014; Faghiri et al., 2014). Building on these studies, we will show that some other commonly accepted asymmetries also turn out to be dubious. These findings lead us to question the almost uncontroversially admitted hierarchical view of the Persian VP and to suggest a flat structure in line with Samvelian (2001) and Bonami & Samvelian (2015). We claim that differences between different types of DOs can be accounted for by semantics, information structure and universal functional principles, such as a “salient-first” preference, without resorting to a hierarchical syntactic structure.

2 The Persian VP: Prevailing Analyses

In (formal) Persian, there is no overt marker for definiteness, ex. *xarguš* ‘the rabbit’. By contrast, indefiniteness is overtly marked by the enclitic *=i*, the cardinal *ye(k)* or both, ex. *yek xarguš=i* ‘a rabbit’. In the DO position, however, NPs with a definite reading are differentially marked by the enclitic *rā*, pronounced (*r*)*o* in the colloquial register, as in ex. (1-a). Moreover, indefinite NPs can be marked by *=rā* to receive an indefinite specific reading as in ex. (1-b).³

- (1) a. Sara *xarguš*(=rā)* did
 Sara rabbit=DOM see.PST.3SG
 ‘Sara saw the rabbit.’
 b. Sara *xarguš=i(=rā)* did
 Sara rabbit=INDEF(=DOM) see.PST.3SG
 ‘Sara saw a (certain/particular) rabbit.’

It should also be noted that in Persian, bare nouns, that is, nouns without any determination or quantification like *xarguš* in (2), are not specified for number and therefore can yield a mass reading. Bare objects have either an existential, ex. (2-a), or a kind-level/generic reading, ex. (2-b). Indefinite objects on the other hand are always specified for number and have an existential reading, as in (1-b) above.

- (2) a. Sara *xarguš* did
 Sara rabbit see.PST.3SG
 ‘Sara saw a rabbit/rabbits.’
 b. Sara *xarguš* dust *dār-ad*
 Sara rabbit friend have.PRS-3SG
 ‘Sara likes rabbits.’

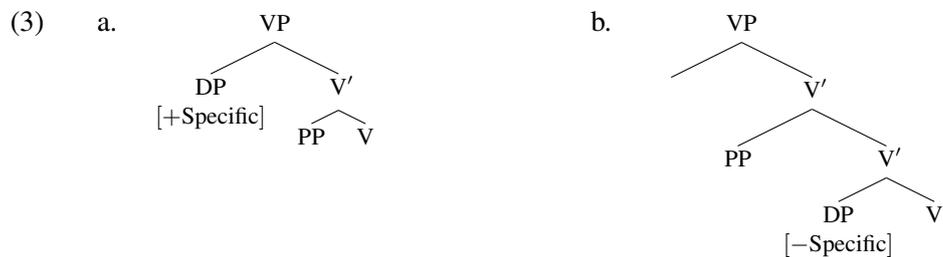
³This explains why many authors have accounted for *=rā* in terms of a binary specificity feature (Karimi, 1990; Browning & Karimi, 1994; Karimi, 2003; Rasekhmahand, 2004), but see Lazard (1982), Dabir-Moghaddam (1992), Meunier & Samvelian (1997) and Ghomeshi (1997).

Note that DOM is a complex phenomenon and that definiteness and/or specificity are not the only features triggering *rā*-marking which can also be triggered by topicality, or more generally, discursive salience. Furthermore, other semantic features such as humanness are also shown to favor the presence of *=rā*.

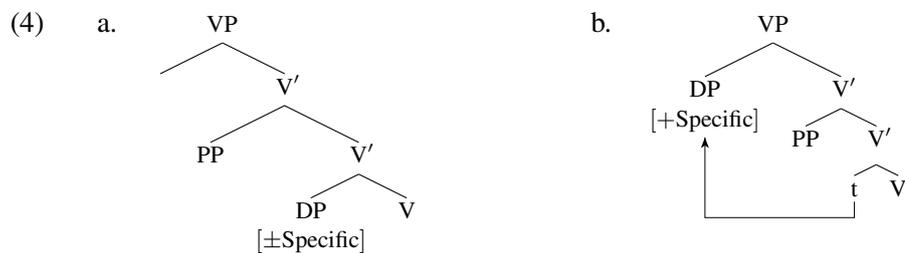
2.1 The “Two Object Position Hypothesis” (TOPH)

Several studies claim that *rā*-marked DOs (definite or indefinite) and unmarked DOs (bare or indefinite) occur in two distinct syntactic positions (at spell out), whether base-generated or as a result of a movement (Karimi, 1990; Browning & Karimi, 1994; Ghomeshi, 1997; Karimi, 2005; Ganjavi, 2007; Modarresi, 2014). Some studies assume different base-generated positions for *rā*-marked and unmarked DOs as in (3-a) and (3-b) respectively, while others assume that both DOs are generated in the same position, as in (4-a), but *rā*-marked DOs must move to the Specifier position to receive interpretation, as illustrated in (4-b).

It should be noted that these studies formulate their claim in terms of features, often a binary one, such as specificity (cf. footnote on page 2), assumed to trigger *rā*-marking, and rarely in terms of DOM itself. However, given the ongoing debate on the analysis of *=rā* and the fact that there is no satisfactory account of *rā*-marking in terms of a binary feature, including specificity, we stick to the formal definition of these two DO types and refrain from referring to any semantic properties.



Adapted from Karimi (2003, p. 105)



Adapted from Karimi (2005, pp. 108–109)

Following Karimi (2003), who provides the most exhaustive argumentation in favor of this analysis, we refer to this claim as the “Two Object Position Hypothesis” (TOPH).

According to Karimi, the fact that *rā*-marked DOs occupy a higher syntactic position than unmarked DOs provides a straightforward account for several syntactic and semantic asymmetries between these two types of DOs. Indeed, the phrase structures in (3) (or in (4)), reflect two different neutral word orders for each type of DO and suggest that only *rā*-marked DOs c-command the indirect object (IO), since unmarked DOs are in a lower position. In the following subsections, we will present these so-called asymmetries. Note however, that we do not agree with a part of Karimi’s grammaticality judgments. The following section provides a discussion of the data on which the TOPH is based.

2.2 The Relative Order with Respect to the IO

According to a widespread hypothesis put forward in theoretical studies as well as (some) grammars, the neutral (unmarked) word order between the direct and indirect objects in ditransitive constructions depends on markedness (Browning & Karimi, 1994; Mahootian, 1997; Rasekhmahand, 2004; Ganjavi, 2007; Windfuhr & Perry, 2009; Roberts et al., 2009, among others). *Rā*-marked DOs are assumed to precede while unmarked DOs follow the IO:

- (5) a. (S) DO=*rā* IO V
 b. (S) IO DO V

The following examples are provided by Karimi (2003) in support of this claim. The author furthermore claims that unmarked DOs can only be separated from the verb if they are contrastively focused.

- (6) a. Kimea aqlab barā mā še’r mi-xun-e
 Kimea often for us poem IPFV-read-3SG
 ‘Kimea often reads poetry for us.’
 b. Kimea aqlab barā mā ye še’r az Hafez mi-xun-e
 Kimea often for us a poem from Hafez IPFV-read-3SG
 ‘Kimea often reads a poem by Hafez for us.’
 c. Kimea aqlab hame=ye še’r-ā=ye tāza=š=**ro** barā mā
 Kimea often all=EZ poem-PL=EZ new=3SG=DOM for us
 mi-xun-e
 IPFV-read-3SG
 ‘Kimea often reads all her new poems for us.’
 d. Kimea aqlab ye še’r az Hafez=**ro** barā mā mi-xun-e
 Kimea often a poem from Hafez=DOM for us IPFV-read-3SG
 ‘Kimea ofte reads a (particular) poem by Hafez for us.’

Adapted from Karimi (2003, p. 91)

2.3 Semantic Fusion with the Verb

Karimi considers *rā*-marked DOs as (independent) participants of the event described by the verb and hence semantically autonomous. Unmarked DOs, by contrast, are assumed to be a part of the predicate and semantically non-autonomous and hence correspond to a [N+Ving] interpretation. This entails that ex. (7), in which the DO is unmarked, can be an appropriate answer to the question ‘What does Kimea do every night?’ (2003, p. 100).

- (7) Kimea har šab (ye) sib mi-xor-e
Kimea every night (a) apple IPFV-eat.PRS-3SG
‘Kimea eats apples / does (an) apple eating every night.’

According to Karimi this difference also accounts for the fact that sentences containing unmarked DOs can only receive an activity/process reading, ex. (8-a), while those containing a marked DO have an eventive reading, ex. (8-b). Note that Karimi borrows this “durative adverbial test” from Ghomeshi & Massam (1994, pp. 190–191), who provide the grammaticality judgments in ex. (8-a) to support the analysis of bare objects (and not for all unmarked DOs) as a case of noun incorporation. Note that Ghomeshi & Massam’s analysis in terms of noun incorporation only applies to bare objects. Indefinite unmarked objects are not concerned by this analysis.

- (8) a. (man) *dar do daqiqe / barāye yek sāat sib xord-am
I in two minute / for one hour apple eat.PST-1SG
‘I ate apples for one hour.’
b. (man) dar do daqiqe / *barāye yek sāat sib=rā xord-am
I in two minute / for one hour apple eat.PST-1SG
‘I ate the apple in two minutes.’

Furthermore, Karimi (2003) claims that the semantic fusion of unmarked DOs with the verb explains why the latter, contrary to marked DOs, cannot take wide scope (and hence cannot trigger scope ambiguity), enter binding relations, and license parasitic gaps. Below, we will present Karimi’s data in support of these claims.

2.3.1 Scope Ambiguity

As illustrated by the pair of examples in (9) and (10), according to Karimi (2003) only *rā*-marked (indefinite) DOs trigger scope ambiguity when scrambled to the left of a DP quantified by a universal quantifier. Note that according to Karimi scope ambiguity can only result from scrambling in Persian (also see Karimi, 2005).

- (9) a. **har** dānešju=i **ye** še'r=ro bāyād be-xun-e
 every student=INDF a poem=DOM must SUBJ-read.PRS-3SG
 'Every student has to read one poem (out of a specific set).' ($\forall > \exists$)
 b. **ye** še'r=ro_i **har** dānešju=i t_i bāyād be-xun-e ($\forall > \exists ; \exists > \forall$)
- (10) a. **har** dānešju=i **ye** še'r bāyād be-xun-e
 every student=INDF a poem must SUBJ-read.PRS-3SG
 'Every student must read a poem.' ($\forall > \exists$)
 b. **ye** še'r_i **har** dānešju=i t_i bāyād be-xun-e ($\forall > \exists$)
 Adapted from Karimi (2003, p. 103)

Karimi argues that unmarked indefinite DOs as part of the predicate can never take wide scope over the IO. That is, only (9-b) allows for both the wide or the narrow scope of existential quantifier over the universal quantifier (2003, p. 103).

2.3.2 Binding Relations

With respect to binding relations, Karimi provides the pair of examples in (11)⁴ in support of the claim that only *rā*-marked DOs are able to bind an anaphor in the IO position.

- (11) a. man [se=tā bačče-hā=ro]_i be hamdige_i mo'arrefi
 I three=CLF child-PL=DOM to each other introduction
 kardam
 do.PST-1SG
 'I introduced three children to each other.'
 b. *man [se=tā bačče]_i be hamdige_i mo'arrefi kardam
 Adapted from Karimi (2003, p. 102)

Karimi (2003) argues that unmarked DOs as part of the predicate cannot enter binding relations.

2.3.3 Licensing Parasitic Gaps

Likewise, with respect to parasitic gaps, Karimi claims that unmarked DOs, contrary to *rā*-marked DOs cannot license parasitic gaps as part of the predicate. This claim is illustrated by the grammaticality judgments provided by the author in (12).

⁴Note that the plural marker, *i.e.* *-hā*, is not compatible with the indefinite NP *se(=tā) bačče* in (b), since this marker implies a definite reading. Accordingly, in the DO position, the plural marker goes in pair with *=rā*-marking.

- (i) se=tā bačče(-hā) xandid-and
 three=CLF child(-PL) laugh.PST-3PL
 '(The) three children laughed.'

- (12) a. Kimea in ketāb=o_i [qablaz in-ke -_i be-xun-e] t_i be
 Kimea this book=DOM before that SUBJ-read.PRS-3SG to
 man dād
 me give.PST.3SG
 ‘Kimea gave me this book before reading (it).’
 b. *Kimea ketāb_i [qablaz in-ke -_i be-xun-e] be man t_i dād
 Adapted from Karimi (2003, p. 116)

2.4 Coordinate Structures

The last argument put forward in support of the TOPH is the claim that *rā*-marked and unmarked DOs cannot appear together in a coordination, as illustrated by Karimi’s grammaticality judgments in (13).

- (13) a. man diruz [in aks=ro] va [in ketāb=ro] xarid-am
 I yesterday this picture=DOM and that book=DOM buy.PST-1SG
 ‘Yesterday, I bought this picture and that book.’
 b. man diruz [aks] va [ketāb] xarid-am
 I yesterday picture and book buy.PST-1SG
 ‘Yesterday, I bought pictures and books.’
 c. *man diruz [in aks=ro] va [ketāb] xarid-am
 I yesterday this picture=DOM and book buy.PST-1SG
 Karimi (2003, p. 103)

According to Karimi (2003) this is a clear argument highlighting the fact that *rā*-marked and unmarked DOs do not occur in the same syntactic position.

3 Getting the Facts Right

The main problem with the TOPH is the fact that the data on which it is built are empirically dubious. The asymmetries either do not hold or are best represented as a cline and certainly not in terms of a dichotomy. More specifically, by adopting a dichotomous view of DOs in terms of *rā*-markedness, this hypothesis aligns indefinite DOs with bare DOs, which are commonly taken to represent all unmarked DOs. This is somewhat expected, albeit misleading, given that bare DOs display the lowest level of specificity and definiteness among unmarked DOs. Yet, bare DOs differ from non-bare unmarked DOs in many non-trivial respects (Ghomeshi & Massam, 1994; Samvelian, 2001; Ghomeshi, 2003; Ganjavi, 2007; Modarresi, 2014).⁵ However, even when these differences are acknowledged and mentioned by scholars, they are put aside and ignored when dealing with the syntactic configuration of the VP (e.g. Ghomeshi, 1997; Ganjavi, 2007).

⁵In the generative frameworks, some scholars have suggested that bare nouns are inserted as N⁰s while indefinites have maximal projections, and are built as NumPs (e.g. Ganjavi, 2007).

In what follows, we will assess the data put forward in the literature, especially in Karimi's studies, in favor on the TOPH.

3.1 Word Order Preferences in Ditransitive Sentences (Revisited)

Recent empirical studies on word order variations in Persian (Faghiri & Samvelian, 2014; Faghiri et al., 2014, 2016) invalidate the generalization in (5) above, clearly showing that unmarked DOs do not display a homogeneous behavior with respect to word order and should be divided into different categories.

Adopting a quantitative approach to word order variations (e.g. Wasow, 1997; Stallings et al., 1998; Yamashita & Chang, 2001; Wasow, 2002; Bresnan et al., 2007), these studies investigate the relative order between the two objects in ditransitive constructions in Persian and examine the effect of functional factors such as the relative length and animacy on ordering preferences. More precisely, Faghiri & Samvelian (2014) explore the predictions of the DOM criterion on the relative order between the DO and the IO (cf. (5)) in a sample of 905 occurrences extracted from the Bijankhan corpus.⁶ Their study is based on a fine-grained typology of DOs with respect to their degree of determination or definiteness: *rā*-marked, indefinite/quantifier (unmarked), bare-modified and bare (single-word) DOs (2014, pp. 222–224). Their data contradict the predictions of the DOM criterion for indefinite (unmarked) DOs, since these DOs occur in the DO-IO-V order at the rate of 77%, grouping with *rā*-marked DOs instead of bare (unmarked) DOs (with 90% and 19.3% of DO-IO-V order respectively). Moreover, interestingly, bare modified DOs, that is, bare DOs carrying modifiers, are shown to have a significantly less stronger preference to appear adjacent to the verb than bare single-word DOs (33.3% vs. 15.8%).

Faghiri et al. (2014) ran a follow up sentence completion (web-based) questionnaire to study the ordering preference of indefinite (unmarked) DOs in a controlled experiment⁷ and arrived at similar results. In their data, the mean rate of sentences like ex. (14) in which the participants placed the DO before the IO (68%) was significantly greater than the opposite order predicted by the DOM criterion.

- (14) ... [DO yek livān (šarbat=e sekanjebin=e tagari)] [IO be moštari-hā (ke
a glass syrup=EZ (mint=EZ icy) to customer-PL (that
az garmā kalāfe bud-and)] dād
from heat frustrated were) gave
'... (he) gave a glass of (icy mint) syrup to the customers (that were frustrated from the heat).'

Adapted from Faghiri et al. (2014, p. 230)

Accordingly, these studies suggest that the DOM criterion should be revisited

⁶A freely available corpus of more than 2.6 million tokens, extracted from the *Hamshahri* daily newspaper contains, manually annotated for part-of-speech information (Faghiri & Samvelian, 2014, p. 222).

⁷For details see Faghiri & Samvelian (2014, pp. 229–232).

in such a way to account for the ordering preferences between the two objects on the basis of the degree of determination of the DO instead of its markedness. In other words, ordering preferences between the DO and the IO rather than being dichotomous, as predicted by the TOPH, follow a continuum based on the degree of determination of the DO : the more determined the DO, the more it is likely to precede the IO. Building on the fact that the degree of determination of the DO is strongly related to its discourse accessibility (cf. e.g. the *(Referential) Givenness Hierarchy*, Gundel et al., 1993), they argue that these ordering preferences reflect a “(discourse-)prominent-first” preference with bareness strongly favoring the IO-DO-V order and *rā*-markedness the inverse.

Furthermore, both Faghiri & Samvelian (2014) and Faghiri et al. (2014) find a significant effect of the relative length corresponding to the “long-before-short” preference. Following Yamashita & Chang (2001), these studies account for this preference in terms of the conceptual accessibility hypothesis. In other words, the “salient-first” preference, which assumes that longer constituents – being lexically richer – are conceptually more accessible than shorter ones (also see Faghiri et al., 2016).

In sum, these quantitative studies show that while *rā*-marked DOs do have a strong preference for the DO-IO-V order, only bare single-word unmarked DOs have a comparable preference to appear adjacent to the verb. Crucially, they show that indefinite (unmarked) DOs group with marked DOs in preferring overall the DO-IO-V order, but show a less strong preference for this order.⁸ Consequently, the empirical findings of Faghiri & Samvelian (2014) and Faghiri et al. (2014) drastically undermine the TOPH, whose backbone argument is the ordering asymmetries between specific (*rā*-marked) DOs on the one hand and unmarked (non-specific) DOs on the other hand.

3.2 Semantic (In)dependence from the Verb (Revisited)

Karimi’s (2003) claim on the semantic fusion of unmarked DOs with the verb faces the same problem as word order preferences, since it similarly builds on the assumption that all unmarked DOs behave in the same way, which is not the case. While bare objects are highly cohesive with the verb, leading some studies to consider them as semantically incorporated, non-bare non-*rā*-marked DOs are inarguably referential NPs and hence are construed as (independent) entities undergoing the event described by the verb rather than being a part of it.

This explains why the “durative adverbial test” argument mentioned by Ghome-shi & Massam (1994) applies only to bare objects. The authors claim that bare DOs are non-referential and as such cannot delimit the event described by the verb and hence are only compatible with adverbials denoting a process. Indefinite unmarked DOs on the other hand are compatible with adverbials denoting an event, ex. (15) .

⁸It should be noted that these differences between the four DO types in their ordering preferences with respect to the IO are replicated in other experiments conducted in the same paradigm as in Faghiri et al. (2014), see Faghiri (2016, ch. 5).

- (15) Maryam dar do daqiqe čand sib=e bozorg xord
 Maryam in two minute a-few apple=EZ big eat.PST.3SG
 ‘Maryam ate a couple of big apples in two minutes.’

3.3 Scope Ambiguity (Revisited)

Karimi’s (2003) judgments and/or interpretations with respect to scope are not straightforward and uncontroversial. In general, scope ambiguity is a complex matter sensitive to functional factors such as discourse and lexical factors. Furthermore, it is shown that scope interpretations display a certain amount of variation among speakers:

Quantifier scope is a delicate phenomenon. It is not simply a matter of ambiguity vs. nonambiguity, but a continuum, and the judgments on a given sentence often fluctuate from speaker to speaker.

Kuno et al. (1999, p. 110)

We believe that any attempt for accounting for scope properties of DOs in Persian is in vain without solid data based on experimental investigations. Meanwhile, crucial to the issue at stake here, not all studies exclude scope ambiguity for unmarked DOs. For instance, Modarresi (2014) assumes both wide and narrow scope for the indefinite DO *film=i* in ex. (16), over the universal quantifier in the subject position. Likewise, Ghomeshi (1997) affirms that the indefinite DO *ye ketāb* in (17) can take both wide and narrow scope over the universal quantifier. Note that the universal quantifier *har* ‘each’, used in Karimi’s (2003) examples above, favors a distributive reading entailing a wide scope, while the universal quantifier *hame* ‘every’ favors a collective reading.

- (16) hame film=i did-and
 everybody movie=INDF watch.PST-3SG
 ‘Everybody watched a movie.’ ($\forall > \exists ; \exists > \forall$)
 Adapted from Modarresi (2014, p. 30)
- (17) hame=ye mo’allem-ā ye ketāb=i entexāb kard-and
 all=EZ teachers-PL a book=INDF selection make.PST-3SG
 ‘All the teachers selected a book.’ ($\forall > \exists ; \exists > \forall$)
 Adapted from Ghomeshi (1997, p. 140)

3.4 Binding Relations (Revisited)

With respect to binding – building on an argument mentioned by Karimi herself (1999, p. 707) –, we have previously suggested, accounting for the ungrammaticality of (11-b) (repeated below in (18-b)) in terms of the semantic mismatch between the pronoun and its antecedent (Faghiri & Samvelian, 2015). Namely, the non-specific NP is not a felicitous antecedent for *hamdige*. Meanwhile, examples like

(19), which are by no means rare as illustrated by the attested example in (20),⁹ show that any generalization based on a binary specific vs. nonspecific feature is too strong. Indeed, these counterexamples show that, contrary to Karimi's claim, an unmarked (that is, nonspecific) DO can bind an anaphor in the IO position.

- (18) a. man [se=tā bačče-hā=ro]_i be hamdige_i mo'arrefi
 I three=CLF child-PL=DOM to each other introduction
 kardam
 do.PST-1SG
 'I introduced three children to each other.'
- b. *man [se=tā bačče]_i be hamdige_i mo'arrefi kardam
- (19) [čand varaq kāqaz]_i be hamdige_i mangane mi-kon-e
 a-few sheet paper to each other staple IPFV-do.PRS-1SG
 '(S)he staples a few sheets of paper together (lit. to each other).'
- (20) ... mi-bin-am [čand=tā sandoq]_i kenār=e hamdige_i
 IPFV-see.PRS-1SG a-few=CLF box next-to=EZ each other
 gozāšt-an ...
 put.PST-3PL
 '... I saw that they have put a few box next to each other...'

A comparison between (19) and (20) on the one hand and (18) on the other hand indicates, in accordance with the line of argumentation pursued here, that plausibly other factors are involved, e.g., humanness of the antecedent, the strength of the distributive reading implied by the predicate (e.g. Dalrymple et al., 1998). Indeed, in the unacceptable example provided by Karimi, ex. (18), contrary to the counterexamples provided here, the referent of the DO is human. Furthermore, in (18), the strong reciprocal predicate ('introduce each one to the other') implies a distributive reading (e.g. Dalrymple et al., 1998), while (19) and (20) are compatible with a collective reading.¹⁰ Interestingly, these factors coincide with those favoring *rā*-marking. As pointed out by several studies, contra Karimi, *rā*-marking is not triggered by a single binary feature. It allows for a certain degree of variability and is sensitive to functional and discourse-related factors (e.g. Samvelian, 2017). The distributive reading implied by the strong reciprocal predicate, the humanness of the referent and the telicity of the event are different converging factors favoring *rā*-marking in (18).

Interestingly, however, the following attested example¹¹ shows that in the proper context, that is, a context favoring a collective reading, even an unmarked DO with

⁹<http://www.noandishaan.com/forums/thread66514.html> [consulted on 07/06/2016]

¹⁰More precisely, (19) favors a collective reading (compare 'staple a few sheets of paper together' with 'staple each sheet to the other'), and (20), while not particularly favoring a collective reading, does not impose a distributive reading in the sense that the sentence does not necessarily require each box to be next to another box.

¹¹<http://romanparisa.blogfa.com/post/10> [consulted on 07/06/2016]

a human referent can bind anaphora in the IO position.¹²

- (21) Lidya yeki=ro mi-šnās-e ke [doxtar pesar]_i [be
Lidya someone=DOM IPFV-know.PRS-3SG that girl boy to
ham_i] mo'arrefi mi-kon-e
each other introduction IPFV-do.PRS-3SG
'Lidya knows someone who introduces boys and girls to each other.'

In any case, further investigation, namely usage-based and experimental studies, is necessary to explore various parameters involved here. Meanwhile, the counterexamples provided here show that the asymmetry claimed by Karimi (2003) on the basis of *rā*-markedness (or specificity for that matter) is flawed.

3.5 Licensing Parasitic Gaps (Revisited)

Contrary to Karimi's claim, unmarked DOs can license parasitic gaps in proper contexts, ex. (22). The problem with Karimi's example in (12-b) (repeated below in (23-b)) results from the incompatibility between the referential properties of *ketāb* and the aspectual properties of the verb in the matrix clause. The telicity of the latter is incompatible with the cumulative reading implied by the bare DO (e.g. Krifka, 1992).

- (22) man tā be hāl na-šode šalvār_i [bedun=e inke -_i porov
I until to now NEG-become.PP pants without=EZ that try
kon-am] -_i be-xar-am
do.PRS-1SG SBJV-buy.PRS-1SG
'It has never happened that I buy pants without trying.'
- (23) a. Kimea in ketāb=o_i [qablaz in-ke -_i be-xun-e] t_i be
Kimea this book=DOM before that SUBJ-read.PRS-3SG to
man dād
me give.PST.3SG
'Kimea gave me this book before reading (it).'
- b. *Kimea ketāb_i [qablaz in-ke -_i be-xun-e] be man t_i dād

Our claim is supported by an acceptability rating experiment, on a Likert scale from 1 (absolutely unacceptable) to 7 (completely acceptable). This experiment was conducted via a web-based questionnaire conducted on the *Ibex Farm* platform (Drummond, 2013),¹³ with 70 Persian native participants. To give an example, the sentence in (22) has received a mean rate of 5.7 (CI_{95%}: ± 0.45).

¹²Note that *ham* is the simplified form of the reciprocal. Indeed, such structures are not fully acceptable with the complex form. But this intolerance has received a functional explanation (cf. Faghiri, 2016, ch. 7).

¹³This platform proposes free hosting for linguistic experiments that can be carried out via on-line questionnaires (<http://spellout.net/ibexfarm>).

Further experiments are needed in order to pin down parameters involved in favoring parasitic gaps with unmarked DOs. In this respect, Goldberg’s discussion on island constraints and the role of the information structure also seems highly relevant (2006, ch. 7). Indeed, the left-extraction of an element needs to be justified on the discursive ground. That is, some discourse saliency is necessary to license the left-extraction of an element. This seems to be the reason why these constructions are more common with *rā*-marked DO, given that *rā*-markedness implies discourse saliency. It should be noted that in Persian the enclitic =*rā* can also be used as a marker of topicality for other non-subject constituents extracted towards the initial position (e.g. Lazard, 1982; Dabir-Moghaddam, 1992).

3.6 Coordination between Marked and Unmarked DOs (Revisited)

Here again, contra Karimi, we claim that the coordination between unmarked and *rā*-marked DOs is perfectly grammatical. Our claim is supported by an acceptability rating experiment, similar to the one presented above, completed by 70 native speakers of Persian via a web-based questionnaire on *Ibex Farm*, with 20 target items and 40 fillers.¹⁴

In this experiment, examples involving a coordination, such as in ex. (25), received an overall mean rate of 5.4 (CI_{95%}: ±0.14), on a scale of 1 to 7, while the control sentence with no coordinate structure, ex. (24), received a mean rate of 6.0 (CI_{95%}: ±0.18). While the difference between the two rates is significant, the mean rate of sentences with a coordinate construction remains high enough to dismiss any doubts on their acceptability.¹⁵

- (24) barāye sabtenām kāfi ast [form=e takmil-šode=rā] barāye mā
 for registration enough is form=EZ completed=DOM for us
 ersāl kon-id
 send do.PRS-2PL
 ‘To register you only need to send us the completed form.’
- (25) a. ... [yek qat’e aks va form=e takmil-šode=rā]
 ... a piece photo and form=EZ completed=DOM
 ‘To register you only need to send us a photo and the completed form.’
- b. ... [form=e takmil-šode=rā va yek qat’e aks]
 ... form=EZ completed=DOM and a piece photo
 ‘To register you only need to send us the completed form and a photo.’

Interestingly, the order between the coordinates is shown to be a relevant fac-

¹⁴Note that for convenience, we are only presenting a simplified version of this experiment here (see Faghiri, 2016, ch. 7, for details)

¹⁵Note that the mean rate of uncontroversially unacceptable sentences, included in the same questionnaire as fillers, is 2.4 (CI_{95%}: ±0.11).

tor. Examples such as (25-a)¹⁶ were rated higher than their counterparts in the reverse order, as in (25-b) : 5.8 (CI_{95%}: ±0.19) vs. 5.0 (CI_{95%}: ±0.20). Further experiments are needed to explore this difference.

4 Less Structure, More Functional/Cognitive Principles

The data presented in the previous section shows that there is no conclusive empirical evidence in favor of the TOPH. If a hierarchical analysis is to be maintained, it should either posit more than two positions, or group unmarked non-bare DOs with *rā*-marked one. None of these solutions is satisfactory, given, among other things, that different types of DOs can be coordinated and that different groupings occur according to the phenomenon considered. Based on this body of evidence, we dismiss this consensual hierarchical analysis and adopt instead a flat structure for the Persian VP, in line with Samvelian (2001) and Bonami & Samvelian (2015).

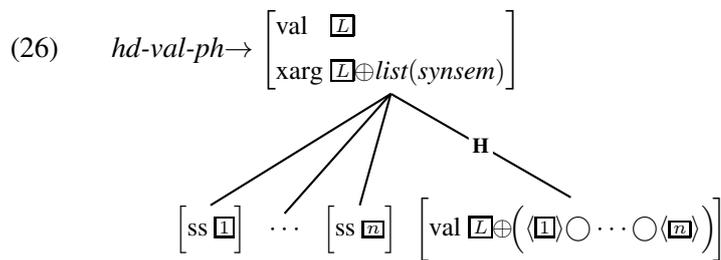
We claim that differences in the behavior of different types of DOs, which constitute a cline rather than a categorical distinction, can be accounted for in terms of universal functional principles. On an unrelated issue, Bonami & Samvelian (2015) claim that word order facts and constituency tests provide no motivation for a VP/S asymmetry in Persian, since subjects and phrasal complements may be freely reordered. We are in line with their view, represented by the head-valence phrase schema given in Figure (26). The schema realizes multiple dependents of the head in the same local tree without constraining their relative order.

Under this view, word order preferences for different DO types can be explained via a set of cross-linguistically valid interacting factors, such as discourse accessibility, definiteness, relative length (or grammatical weight) and animacy, and stated in terms of the principle of “prominent-first”, pointed out for other SOV languages, such as Japanese (Yamashita & Chang, 2001).

¹⁶One can consider these examples misleading for the issue at stake, since the enclitic =*rā* is placed on the right periphery and can theoretically scope over both DOs. To avoid this confusion, we have prepared our stimuli in a way that in each item the indefinite DO would not be (semantically) felicitous if *rā*-marked. In other words, the *rā*-marked variant of the indefinite DO in these sentences yields an awkward utterance:

- (i) ??barāye sabtenām kāfi ast [yek qat'e akse=rā] barāye mā ersāl kon-id
 for registration enough is a piece photo=DOM for us send do.PRS-2PL
 ‘To register you only need to send us a particular photo.’

Nonetheless, the mean rate of (b) sentences, 5.0 (CI_{95%}: ±0.20), remains high enough, since it is significantly higher than 4 (t=9.71, df=279). Recall that the mean rate of uncontroversially unacceptable sentences in the same questionnaire is 2.4 (CI_{95%}: ±0.11).



5 Conclusion

In this paper we showed that the behavior of the DOs in Persian cannot be accounted for in terms of a hierarchical phrase structure, since the differences between different types of DOs are a matter of cline rather than a dichotomous opposition. Trying to account for these empirical facts by adding more structure, as theoretically appealing as it may seem, not only does not provide an appropriate modeling of data but also makes wrong predictions. In contrast, a simplified structure accompanied by few functional principles constitutes a more satisfying option to explore.

References

- Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2). 327–382.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Royal Netherlands Academy of Science.
- Browning, Marguerite & Ezat Karimi. 1994. Scrambling to object position in Persian. In N. Corver & H. van Riemsdijk (eds.), *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*, 61–100. Mouton de Gruyter.
- Dabir-Moghaddam, Mohammad. 1992. On the (in) dependence of syntax and pragmatics: Evidence from the postposition-ra in Persian. *Cooperating with written texts*, 549–573.
- Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo & Stanley Peters. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21(2). 159–210.
- Diesing, Molly. 1992. *Indefinites*. Linguistic Inquiry Monographs. Cambridge, Mass: MIT Press.
- Drummond, Alan. 2013. Ibex farm. Version 0.3. www.spellout.net/ibexfarm/.

- Faghiri, Pegah. 2016. *La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique*, PhD dissertation, Université Sorbonne Nouvelle .
- Faghiri, Pegah, Barbara Hemforth & Pollet Samvelian. 2016. Length effects in an OV language with differential object marking and mixed head direction (poster). In *The 29th annual conference on human sentence processing (CUNY 2016)*, University of Florida, Gainesville, Florida, USA.
- Faghiri, Pegah & Pollet Samvelian. 2014. Constituent ordering in Persian and the weight factor. In Christopher Pinon (ed.), *Empirical issues in syntax and semantics 10 (EISS 10)*, 215–232. CNRS.
- Faghiri, Pegah & Pollet Samvelian. 2015. How much is determined by the syntax? an empirical approach to the position of the direct object in Persian. In *Proceedings of the 9th Iranian conference on linguistics, Allameh Tabataba'i University, Tehran, feb. 24-25 2015*, 1419–1435. Allameh Tabataba'i University.
- Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2014. Accessibility and word order: The case of ditransitive constructions in Persian. In Stefan Müller (ed.), *Proceedings of HPSG 2014*, 217–237. Stanford: CSLI Publications.
- Ganjavi, Shadi. 2007. *Direct objects in Persian*. PhD dissertation, University of Southern California.
- 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.
- Ghomeshi, Jila & Diane Massam. 1994. Lexical/syntactic relations without projection. *Linguistic Analysis* 24(3-4). 175–217.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Gundel, Jeanette K, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.
- Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions: Ra in Persian. *Linguistic Analysis* 20. 139–191.
- Karimi, Simin. 1999. A note on parasitic gaps and specificity. *Linguistic Inquiry* 30(4). 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 Ltd.

- Karimi, Simin. 2005. *A minimalist approach to scrambling: Evidence from Persian*. Studies in Generative Grammar. Mouton De Gruyter.
- Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical matters*, 29–53. Stanford: CSLI Publications.
- Kuno, Susumu, Ken-ichi Takami & Yuru Wu. 1999. Quantifier scope in English, Chinese, and Japanese. *Language* 75(1). 63–111.
- Lazard, Gilbert. 1982. Le morphème *rā* en persan et les relations actanciennes. *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.
- Modarresi, Fereshteh. 2014. *Bare nouns in Persian: Interpretation, grammar and prosody*, PhD dissertation, Humboldt-Universität zu Berlin.
- Rasekhmahand, Mohammad. 2004. *Jāygāh maf'ul mostaqim dar fārsi* [the position of the direct object in Persian]. *Nāme farhangestān* 6(4). 56–66.
- Roberts, John R, Behrooz Barjasteh Delforooz & Carina Jahani. 2009. *A study of Persian discourse structure* Studia Iranica Upsaliensia, 12. 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. to appear. Specific features of the Persian syntax the Ezāfe Construction, Differential Object Marking and Complex Predicates. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *Oxford handbook of persian linguistics*. Oxford: Oxford University Press.
- Stallings, Lynne M, Maryellen C MacDonald & Padraig G O'Seaghda. 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. 1997. Remarks on grammatical weight. *Language Variation and Change* 9(1). 81–105.
- Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: CSLI Publications.
- Windfuhr, Gernot L & John R Perry. 2009. Persian and Tajik. In Gernot L Windfuhr (ed.), *The iranian languages* Routledge Language Family Series, 416–544. London: Routledge.

Yamashita, Hiroko & Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81(2). B45–B55.