

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

Pegah Faghiri & Pollet Samvelian
Université Sorbonne Nouvelle - Paris 3

In this paper we address the issue of the syntactic structure of the VP in Persian, a null subject SOV language with relatively free word order. The prevailing view of the Persian VP initially suggested in generative studies assumes a hierarchical structure, mainly motivated by the existence of differential object marking (DOM), which requires that definite DOs be marked by the enclitic *rā*. It has been claimed that *rā*-marked and unmarked DOs display several syntactic and semantic asymmetries, which can be straightforwardly accounted for if two distinct configurational positions are posited for each type of DO. Following insights from studies such as Diesing (1992), unmarked DOs have been assumed to be VP internal while *rā*-marked DOs are VP external (Karimi, 1990; Browning and Karimi, 1994; Ghomeshi, 1997; Karimi, 2003; Ganjavi, 2007). The higher syntactic position of the latter explains word order preferences in ditransitive 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 and 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 and 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 “given-first” preference.

1. The Persian VP : Prevailing Analyses

In (formal) Persian, there is no overt marker for definiteness. By contrast, indefiniteness is overtly marked by the enclitic *=i*, the cardinal *ye(k)* or both. Furthermore, bare nouns, i.e. nouns without any determination or quantification like *xarguš* in (1), are not specified for number and therefore can yield a mass reading: they have either an existential or a kind-level/generic reading. Indefinite NPs on the other hand are always specified for number and have an existential reading, ex. (2) and (3). Definite DOs are necessarily marked by the enclitic *rā* (colloquial: *(r)o*), (4). Meanwhile, DOM is not incompatible with the indefinite determination, ex. (3). In this case, the NP receives an indefinite specific reading.

- | | | | |
|-----|--|-----|--|
| (1) | Sara xarguš did
Sara rabbit saw
'Sara saw a rabbit/rabbits.' | (3) | Sara (yek) xarguš= <i>i=rā</i> did
Sara (a) rabbit=INDEF=DOM saw
'Sara saw a (certain/particular) rabbit.' |
| (2) | Sara (yek) xarguš= <i>i</i> did
Sara (a) rabbit=INDEF saw
'Sara saw a rabbit.' | (4) | Sara xarguš= <i>rā</i> did
Sara rabbit=DOM saw
'Sara saw the rabbit.' |

The “Two Object Position Hypothesis” (TOPH, Karimi, 2003) claims that *rā*-marked (specific in Karimi’s terms) DOs (definite or indefinite), on one hand, and unmarked (nonspecific) DOs (bare or indefinite), on the other hand, occur in two different syntactic positions. Figures (12-a) and (12-b) from Karimi (2003) illustrate these two positions and are assumed to account for the asymmetries presented in (A)-(F) below. Note that we do not always agree with Karimi’s grammaticality judgments (cf. section 2).

A. Word Order: In ditransitive constructions, *rā*-marked DOs precede while unmarked DOs follow the IO:

- | | | | |
|-----|----|---|------------------------------------|
| (5) | a. | Kimea aqlab barā mā še'r mi-xun-e
Kimea often for us poem IPFV-read-3SG
'It is often the case that Kimea reads poetry for us.' | |
| | b. | Kimea aqlab hame=ye še'r-ā=ye tāza=š=ro barā mā mi-xun-e
Kimea often all=EZ poem-PL=EZ new=3SG=DOM for us IPFV-read-3SG
'It is often the case that Kimea reads all her new poems for us.' | Adapted from Karimi (2003, p. 91). |

B. Semantic Fusion with the Verb: *Rā*-marked DOs are considered as (independent) participants of the event described by the verb and hence semantically autonomous, while unmarked DOs are assumed to be a part of the predicate, and semantically non autonomous. This entails that (6), where the DO is unmarked, can be an appropriate answer to the question ‘What does Kimea do about...?’. With a *rā*-marked DO, the same sentence could only be an answer to the question ‘What does Kimea do to an/the apple every night?’ Karimi (2003, p. 100). This difference also accounts for the fact that sentences containing unmarked DOs can only receive an activity/process reading, ex. (7-a), while those containing a marked DO have an eventive reading, ex. (7-b).

- | | |
|-----|--|
| (6) | 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.' |
|-----|--|

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

Adapted from (cf. Ghomeshi and Massam, 1994, pp. 190-191)

C. Scope Ambiguity: Only *rā*-marked DOs trigger scope ambiguity when scrambled left to a DP quantified by a universal quantifier. That is, in (8), only (8-b) allows for both the wide or the narrow scope of existential quantifier over the universal quantifier (Karimi, 2003, p. 103).

- (8) 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 ($\forall > \exists ; \exists > \forall$)
- c. **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.' ($\forall > \exists$)
- d. **ye** še'r_i **har** dānešju=i t_i bāyād be-xun-e ($\forall > \exists$)

D. Binding. Only *rā*-marked DOs are able to bind an anaphora in the IO position (Karimi, 2003, p. 102):

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

E. Licensing Parasitic Gaps. Only *rā*-marked DOs can license parasitic gaps (Karimi, 2003, p. 116) :

- (10) a. Kimea in ketāb=o_i [qablaz in-ke -_i be-xun-e] t_i be man dād
Kimea this book=DOM before that SUBJ-read.PRS-3SG to 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

F. Coordination. Specific and nonspecific DOs can not appear together in a coordination (Karimi, 2003, p. 103):

- (11) a. man diruz [in aks=ro] va [in ketāb=ro] xarid-am
I yesterday this picture=DOM and that book=DOM buy.PST-1 SG
'Yesterday, I bought this picture and that book.'
- b. man diruz [aks] va [ketāb] xarid-am
I yesterday picture and book buy.PST-1 SG
'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-1 SG

2. Getting the facts right

The main problem with the TOPH is the fact that the data on which it builds on is empirically dubious. The asymmetries either do not hold or are best represented as a cline and certainly not in terms of dichotomy. More specifically, by adopting a dichotomous view of DOs in terms of *rā*-markedness, this hypothesis aligns indefinite DOs with bare DOs. The latter are commonly taken to represent all unmarked DOs (bare and indefinite), which is somewhat expected, albeit misleading, given that bare DOs display the lowest level of specificity and definiteness. Yet, they differ from non-bare unmarked DOs, that is, DOs carrying an indefinite determination, in many non-trivial respects (e.g. Ghomeshi and Massam, 1994; Samvelian, 2001; Ghomeshi, 2003; Ganjavi, 2007; Modarresi, 2014).

A. Canonical word Order and the behavior of non-specific DOs: Recent empirical studies on word order variations in Persian (Faghiri and Samvelian, 2014; Faghiri et al., 2014), including corpus and experimental data, clearly show that non-specific DOs do not display a homogenous behavior in this respect. Crucially, while bare objects indeed follow the IO and thus appear adjacent to verb, indefinite unmarked (i.e. non specific) DOs group with specific (*rā*-marked) DOs and precede the IO. Based on a fine-grained typology of Persian NPs with respect to their degree of determination, the authors suggest that ordering preferences between the DO and the IO follow a continuum with bareness strongly favoring the IO-DO-V order and *rā*-markedness the inverse, and are influenced by functional factors such as relative length (corresponding to the "long-before-short" preference). The empirical findings of Faghiri and Samvelian (2014); Faghiri et al. (2014) drastically undermine the TOPH, whose backbone argument is the ordering asymmetries between specific (*rā*-marked DOs) on one hand and unmarked (non-specific) DOs on the other hand.

B. Semantic (in)dependence from the verb: Karimi’s claim on the semantic fusion of nonspecific DOs with the verb faces the same problem as word order preferences, since it similarly builds on the assumption that nonspecific 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 Ghomeshi and 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 being referential are compatible with adverbials denoting an event.

C. Licensing Parasitic Gaps: Contrary to Karimi’s claim, unmarked DOs do license parasitic gaps, ex. (12). This claim is supported by our preliminary acceptability rating experience. Further experiences are currently carried on in order to pin down parameters involved in favoring parasitic gaps with unmarked DOs. The problem with Karimi’s example in (10) 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 (cf. e.g. Krifka, 1992).

- (12) man bastani_i [bā-inke xeyli –_i dust dār-am] ba’d-az šām –_i ne-mi-xor-am
 I ice-cream even-though very like have.PRS-1SG after dinner NEG-IPFV-eat.PRS-1SG
 ‘I don’t eat ice cream after dinner even though I like (ice cream) very much.’

D. Coordination between marked and unmarked DOs: Here again, contra Karimi and on the basis of acceptability rating experiences, we claim that the coordination between unmarked and *rā*-marked DOs is perfectly grammatical, as illustrated by (13).

- (13) a. man diruz [čand-tā ruznāme] va [in majalla=ro] xarid-am
 I yesterday a-few newspaper and this magazine=DOM buy.PST-1SG
 ‘Yesterday, I bought a few newspapers and this magazine.’
 b. man diruz [ruznāme], [čand-tā majalla] va [in ketāb=ro] xarid-am
 I yesterday newspaper a-few magazine and this book=DOM buy.PST-1SG
 ‘Yesterday, I bought a newspaper/newspapers, some magazines and this book.’

3. 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 with this consensual hierarchical analysis and adopt instead a flat structure for the Persian, in line with Samvelian (2001) and Bonami and 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 and 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 stick to their view, represented by the head-valence phrase schema given in Figure (16). The latter 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, 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 and Chang, 2001).

With respect to binding, we have previously suggested (Faghiri and Samvelian, 2015) – building on an argument mentioned by Karimi herself 1999, p. 707 –, to account for the ungrammaticality of (9) in terms of the semantic mismatch between the pronoun and its antecedent. Namely, the nonspecific NP is not a felicitous antecedent for *hamdige*. Meanwhile, examples like (14) show that any generalization based on a binary specific vs. nonspecific feature is too strong and therefore flawed. Indeed, in a proper context a nonspecific DO can bind the IO.

- (14) [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).’

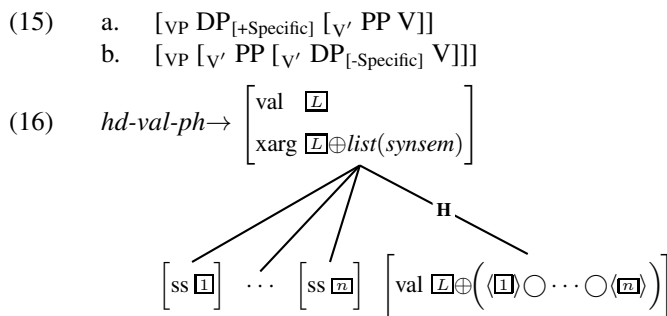
A comparison between (14) and (9) indicates, in consistency with the line of argumentation pursued here, that plausibly other factors play a role, e.g., humanness of the antecedent, the strength of the distributive reading implied by the predicate (Dalrymple et al., 1998). It is interesting to note that this is reminiscent of *rā*-marking in Persian. 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 (cf. e.g. Samvelian, 2016). In any case, further investigation is necessary to identify various parameters involved.

The same holds for scope relations. Not all studies accept Karimi’s acceptability judgments and generalizations. For instance, Modarresi and Simonenko (2007) and Modarresi (2014) claim that indefinite unmarked DOs can display both narrow and wide scope over a universal quantifier. Any attempt for accounting for scope properties of DOs in Persian is vain without solid data based on experimental investigations.

4. 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. On the contrary, a simplified structure accompanied by few functional principles constitutes a more satisfying option to explore.

Figures



References

- Bonami, O. and Samvelian, P. (2015). The diversity of inflectional periphrasis in Persian. *Journal of Linguistics*, 51(02):327–382.
- Browning, M. and Karimi, E. (1994). Scrambling to object position in Persian. *Studies in scrambling*, pages 61–100.
- Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., and Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy*, 21(2):159–210.
- Diesing, M. (1992). *Indefinites*. Linguistic Inquiry Monographs. Cambridge, Mass: MIT Press.
- Faghiri, P. and Samvelian, P. (2014). Constituent ordering in Persian and the weight factor. In Pinon, C., editor, *Empirical Issues in Syntax and Semantics 10 (EISS 10)*, pages 215–232. CNRS.
- Faghiri, P. and Samvelian, P. (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*, pages 1419–1435. Allameh Tabataba’i University.
- Faghiri, P., Samvelian, P., and Hemforth, B. (2014). Accessibility and word order: The case of ditransitive constructions in Persian. In Müller, S., editor, *Proceedings of HPSG 2014*, pages 217–237. Stanford: CSLI Publications.
- Ganjavi, S. (2007). *Direct Objects in Persian*. University of Southern California.
- Ghomeshi, J. (1997). Topics in Persian VPs. *Lingua*, 102(2):133 – 167.
- Ghomeshi, J. (2003). Plural marking, indefiniteness, and the noun phrase. *Studia linguistica*, 57(2):47–74.
- Ghomeshi, J. and Massam, D. (1994). Lexical/syntactic relations without projection. *Linguistic Analysis*, 24(3-4):175–217.
- Karimi, S. (1990). Obliqueness, specificity, and discourse functions: Ra in Persian. *Linguistic Analysis*, 20:139–191.
- Karimi, S. (1999). A note on parasitic gaps and specificity. *Linguistic Inquiry*, 30(4):704–713.
- Karimi, S. (2003). On object positions, specificity, and scrambling in Persian. In Karimi, S., editor, *Word Order and Scrambling*, pages 91–124. Blackwell Publishing Ltd.
- Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In Sag, I. A. and Szabolcsi, A., editors, *Lexical matters*, pages 29–53. Stanford University.

- Modarresi, F. (2014). *Bare nouns in Persian: Interpretation, Grammar and Prosody*. PhD thesis, Humboldt-Universität zu Berlin.
- Modarresi, F. and Simonenko, A. (2007). Quasi noun incorporation in Persian. In *LingO*, pages 181–189.
- Samvelian, P. (2001). Le statut syntaxique des objets nus en persan. *Bulletin de la Société de Linguistique de Paris*, 96(1):349–388.
- Samvelian, P. (2016). Specific features of the Persian syntax the Ezāfe Construction, Differential Object Marking and Complex Predicates. In *Handbook of Persian*, page to appear. Oxford Univ Press.
- Yamashita, H. and Chang, F. (2001). “Long before short” preference in the production of a head-final language. *Cognition*, 81(2):B45–B55.