The issue of ”separability” in Persian complex predicates

Pegah Faghiri, Pollet Samvelian

Chapter 1

The issue of "separability" in Persian complex predicates

Pegah Faghiri
Universität zu Köln

Pollet Samvelian
Université Sorbonne Nouvelle

This paper addresses the issue of separability in Persian complex predicates (CPs). These are syntactic combinations formed by a verb and a preverbal element (noun, adjective, preposition) realizing a single conceptual unit. Although the separability of the components of a CP by morphological and grammaticalized elements (ex. auxiliaries) is not a matter of controversy, the possibility for "real" syntactic constituents to interrupt a CP continues to be debated. Building on an experiment-based study, we show that real syntactic material can separate the components of a CP and suggest that this separability can be viewed as a word order variation phenomenon, comparable to the one observed for direct objects (DO) and indirect objects (IO) in the preverbal domain. The semantic bond plays nevertheless a role in granting CPs some hallmarks of "wordhood", favoring their adjacency, among other things.

1 Introduction

In this paper, we address the issue of separability in Persian complex predicates (CPs). Building on an experiment-based study, we show that real syntactic material can separate the components of a CP, a possibility generally underestimated or denied in most previous studies on Persian CPs. We also suggest that this separability can be viewed as a word order variation phenomenon, comparable to the one observed for direct objects (DO) and indirect objects (IO) in the preverbal domain. As such, it is best accounted for by soft constraints on word order,

that is (statistical) preferences, involving a set of functional factors, rather than by categoric syntactic, or phrase structure, rules (hard constraints). Likewise, we do not consider that the strong preference for the components of the CP to occur adjacent to each other is peculiar to CPs, hence requiring a specific syntactic treatment. This preference is also observed for bare objects in Persian, which tend to occur adjacent to the verb. The fact that it becomes even stronger in the case of CPs is totally expected, given that semantic relatedness favors adjacency. Thus, on the one hand, the fact that several words form a single conceptual unit favors their remaining together (one semantic unit), while on the other hand, the fact that the sequence is made of multiple syntactic units still allows for the word order preference rules to apply.

It is a well known fact that the verbal lexicon in Persian is overwhelmingly formed by complex predicates, that is, multitword expressions including a verb and a non-verbal element, mainly a noun, such as *bāzi kardan* 'to play' (play do) or *qadam zadan* 'to walk' (step hit), also known as "light verb constructions" (LVCs).[1]

Forming one semantic unit, the components of a CP tend to remain together and resist separation, except by morphological or grammaticalized material (verbal prefixes, clitic pronouns, auxiliaries). This has led many studies to take a strong stance on this issue, claiming that "real" syntactic material can never intervene between the verb and the non-verbal unit of a CP. This claim has served as a key argument in favor of the "wordhood" (Goldberg 1996: 134–135) or a "lexical analysis" (Dabir-Moghaddam 1997; Karimi-Doostan 1997) of CPs, along with other properties, which are typical of words, or rather lexemes in this case. Namely:

- The whole sequence generally has a conventional meaning that must be learned by the speakers. In other words, it is idiomatic, in that the meaning associated to the sequence cannot be totally derived from its components' meaning (Goldberg 1996; Karimi-Doostan 1997; Samvelian 2001; 2012; Samvelian & Faghiri 2013).

- It can serve as input to morphological word formation rules that derive new lexemes from existing ones (Goldberg 1996; Karimi-Doostan 1997; Megerdoomian 2002; Vahedi-Langrudi 1996).

[1]There are also CPs formed with an adjective, ex. *bāz kardan* 'to open' (open do), a preposition or particle, ex. *bar dāštan* 'to take' (PART have) or a prepositional phrase *be kār bordan* 'to use' (to work take). In this paper, we will focus on noun-verb CPs.

The supposed inseparability of the CP components has further been used to draw a clearcut distinction between the latter on the one hand and ordinary verb-complement syntactic combinations on the other hand, and to support a specific syntactic analysis of CPs, which distinguishes them from ordinary syntactic combinations involving a verb and its object (with the notable exception of Müller 2010; Samvelian 2001; 2012; Samvelian & Faghiri 2014; 2016).

Although Samvelian (2012: 55–87) extensively discusses this issue and provides several attested examples showing that almost all CPs can undergo separation, the controversy seems still persisting since more recent studies (Safavi et al. 2016: among others) take the inseparability of at least some classes of CPs as empirically uncontroversial.

In this paper, we will first present the basic empirical facts about Persian CPs and their syntactic properties as they have been discussed in the literature, with a special focus on the issue of separability. In particular, we will examine Karimi-Doostan's claim about the relationship between the separability and the predicative nature of the nominal element in Noun-Verb CPs. Contra Karimi-Doostan, we will provide an experimental-based study which shows that the nominal element of a CP, regardless of its type and its degree of determination, can be separated from the verb by syntactic material.

Comparing the results of our experiments with the findings of some recent studies on word order variations in the preverbal domain in Persian (Faghiri 2016; Faghiri & Samvelian 2014; Faghiri et al. 2018), which also resort to quantitative methods, we will argue that Noun-Verb CPs, on the whole, behave in the same way as DO-verb combinations with respect to word order preferences. Crucially, the latter involves preferences rather than strict syntactic constraints.

It has been shown that different (functional) factors (ex. givenness, animacy, length) interact to determine the linear order of constituents, where the latter is not constrained by the grammar. Some of these factors (degree of determination, heaviness and animacy) has also been shown to intervene in ordering preferences regarding direct and indirect objects in Persian as well (Faghiri 2016). We will see that the same factors are at play in determining the ordering preferences of CPs components. Furthermore, semantic relatedness and collocational relation are two factors known to favor adjacency (see Hawkins 2001; Wasow 2002: among others). Hence, the tendency for the components of a CP to appear adjacent is not surprising, since they convey one conceptual meaning.

2 Existing claims on the inseparability of CPs

Several studies on Persian CPs claim that the separability of the components of a CP is subject to significant restrictions. According to Goldberg (1996), only morphological and "grammatical" material may intervene between the non-verbal element and the verb, as illustrated in (1).[2]

(1) a. omid goli=rā setāyeš ne-mi-konad
Omid Goli=RA praise NEG-IPFV-do.PRS-3SG
'Omid doesn't praise Goli.'

b. omid setāyeš=aš kard
Omid praise=CL.3SG did.PST
'Omid praise her/him.

c. omid setāyeš=aš xāhad kard
Omid praise=CL.3SG AUX.FUT.3SG do.SINF
'Omid will praise her/him.'

In (1a), the nominal element of the CP *setāyeš kardan* 'to praise' (praise do), namely *setāyeš* 'praise', is separated from the verb by the negation prefix *na-* and the aspect-mode prefix *mi-*; in (1b), the clitic pronoun *=aš*, which refers to the direct object in the first example, attaches to the nominal element and thus separates it from the verb. Finally, in (1c), the intervening element is the tense auxiliary *xāstan* 'to want', which is an independent word.

"Real" syntactic material, on the other hand, Goldberg (1996) claims, cannot occur between the components of the CP. This restriction is illustrated by examples in (2), adapted from Goldberg (1996: 134–135):

(2) a. tond rānandegi kard-am
quickly driving do.PST-1SG
'I drove quickly.'

b. ?? rānandegi tond kard-am
driving quickly do.PST-1SG
(Intended) 'I drove quickly.'

[2]Glosses follow the Leipzig Glossing Rules (www.eva.mpg.de/lingua/resources/glossing-rules.php). The following non-standard abbreviations are used for clarity: RA = differential object marker; EZ = Ezafe.

(3) a. ali=rā setāyeš kard-am
Ali=RA praise do.PST-1SG
'I adored Ali.'

b. ?? setāyeš ali=ā kard-am
praise Ali=RA do.PST-1SG
(Intended) 'I adored Ali.' (Goldberg 1996, p. 135, ex. 3)

According to Goldberg (1996), (2b) shows that the placement of a modifier adverb, *tond* 'quickly' here, between the nominal element and the verb makes the sentence odd. The adverb must precede the whole CP, as in (2a). While, in ordinary object-verb combinations, a modifier adverb can intervene between the object and the verb, as shown by (4). Example (3b) shows that the direct object cannot interrupt the CP and must be placed before it, as in (3a).

(4) mašq=am=rā tond nevešt-am
homework=CL.1SG=RA quickly write.PST-1SG
'I did my homework quickly'(Goldberg 1996, p. 134, ex. 10)

These facts, argues Goldberg (1996), imply that CPs are single syntactically integrated predicates, comparable to some extent to words (or lexical units). As such, they undergo constraints which do not apply for ordinary syntactic combinations. These constraints may nevertheless be violated in some contexts, allowing for morphological (affixes and clitics) and grammatical elements (auxiliaries) to intervene between the components of a CP.

Contrary to Goldberg (1996), Karimi-Doostan (1997, 2011) admits that the components of a CP can be separated by syntactic elements depending on the type of the nominal element of the CP. The latter are classified into three categories: predicative nouns, ex. *latme* 'damage', verbal nouns, ex. *ersāl* 'sending', and non-predicative nouns, ex. *guš* 'ear'. It is claimed that only CPs formed by predicative nouns are separable. The rationale is that for the nominal element to be separable from the verb, it needs to meet the following two conditions (in the context of a given CP):

1. It must have an argument structure.
2. It must be able to project a DP/NP, that is, be determined or quantified.

Only predicative nouns, it is claimed, can fit these conditions, as illustrated by examples (5)–(7).

(5) a. latme=ye tagarg be bāq=e man
damage=EZ hail to garden=EZ I
'hail's damage to my garden'

b. tagarg be bāq=e man latme zad
hail to garden=EZ I damage hit.PST.3SG
'The hail damaged my garden.'

c. tagarg latme=ye bad=i be bāq=e man zad
hail damage=EZ bad=INDF to garden=EZ I hit.PST.3SG
'The hail damaged my garden badly.'

(6) a. anjām-e kār tavassot=e ali
performing=EZ work by=EZ Ali
'Ali's doing the work'

b. ali kār=rā anjām dād
Ali work=RA performing give.PST.3SG
'Ali did the work.'

c. * ali anjām-e xub=i be kār dād
Ali performing=EZ good=INDF to work give.PST.3SG
(Intended) 'Ali did the work well.'

(7) a. * guš=e Ali be rādyo
ear=EZ Ali to radio
(Intended) 'Ali's listening to the radio'

b. ali be rādyo guš dād
Ali to radio ear give.PST.3SG
'Ali listened to the radio.'

c. * ali guš-e xub=i be rādyo dād
Ali ear=EZ good=INDF to radio give.PST.3SG
(Intended) 'Ali listened to the radio well.'

Latme 'damage', ex. (5), is a predicative noun. It has an argument structure, as shown by its ability to realize its arguments within an DP/NP, ex. (5a). As the nominal element of the CP *latme zadan* 'to damage', *latme* must be adjacent to the verb when it is realized as a bare noun, ex. (5b). When determined, the nominal element of the CP functions as the nominal argument of the verb. It becomes autonomous and can be separated from the verb by various syntactic

constituents. This is illustrated by (5c), where *latme* 'damage' carries the indefinite determiner, the enclitic *=i*, and consequently can precede the prepositional argument.

Like predicative nouns, verbal nouns, e.g. *ersāl* 'sending' and *anjām* 'accomplishment, performance', also carry an argument structure, ex. (6a). However, unlike the former, they cannot project a DP/NP, since they have limited nominal behavior: they cannot be pluralized, modified, quantified and determined. These nouns are broadly assumed to form prototypical light verb constructions, ex. *ersāl kardan* 'to send', *anjām dādan* 'to accomplish, to do'. In this case, they always occur in their bare form and hence adjacent to the verb, ex. (6b). These properties of verbal nouns explain the ungrammaticality of (6c).

Finally, non-predicative nouns, ex. *guš* 'ear', do not carry argument structure, as illustrated by (7a). When used outside a CP, these nouns can develop into DP/NPs, ex. *in guš* 'this ear'. However, when used as the nominal element of a CP, ex. *guš kardan* 'to listen', they can only appear in their bare form, ex. (7b), and therefore must remain adjacent to the verb, hence the ungrammaticality of (7c).

3 Severing separability from DP/NP projection

Before investigating the separability of the components of a CP, it should be made clear that Karimi-Doostan's claims involve two different, though perhaps interrelated, issues:

1. The first issue concerns the possibility for the bare nominal element of the CP to be separated from the verb by syntactic material.

2. The second issue is the possibility for the nominal element of the CP to project a DP/NP and thus to behave as an autonomous syntactic constituent with respect to the verb.

Under Karimi-Doostan's view, these two issued are entangled since separation is possible only for DP/NPs. However, several studies on Persian CPs provide examples of bare nominal elements of CPs which are not adjacent to the verb:

(8) a. … va sili be surat=am zad
and slap to face=CL.1SG hit.PST.3SG
'…and (s)he slapped me (on the face).'[3] (Samvelian 2012: p. 40, ex. 29)

[3]This is an attested example taken from the novel *Souvašun* by S. Danešvar (Samvelian 2012).

b. guš be man ne-mi-kon-e
ear to me NEG-IPFV-do-3SG
'(S)he doesn't listen to me.' (Mohammad & Karimi 1992: p. 197, ex. 7)

c. kimiā in otāq=rā extesās be mehmān dād
Kimea this room=RA allocation to guest give.PST.3SG
'Kimea allocated this room to the guest.' (Mohammad & Karimi 1992: p. 199, ex. 16)

In (8a), the predicative noun *sili* 'slap', which occurs as a bare noun, is nevertheless separated from the verb by the PP argument of the CP. In (8b), the PP argument intervenes between the non-predictive noun *guš* 'ear', again in its bare form, and the verb. (8c) illustrates the possibility for a verbal noun to precede the PP argument.[4]

These examples show that the possibility for bare nominal elements of CPs to be separated from the verb is a matter of controversy. Contra Goldberg (1996) and Karimi-Doostan (1997; 2011), Samvelian (2012) claims that the adjacency of the bare nominal element and the verb in a CP is a matter of strong preference and not a strict constraint. She further draws a parallel between these bare nominal elements and bare objects of lexical verbs, which also tend to occur adjacent to the verb, as it has been noted in all studies on the syntax of Persian (Dabir-Moghaddam 1997; Givi Ahmadi & Anvari 1995; Ghomeshi 1996; Lazard 1982; Mahootian 1997; Samvelian 2001; Karimi 2003: among many others). Like bare objects, bare nominal elements of CPs can nevertheless be separeted from the verb by syntactic material. Their greater reluctance to separation, compared to bare objects of lexical verbs, is due to the idiomatic relation between the components of a CP and their closer semantic relatedness, which favors even more adjacency.

To sum up, one issue to be addressed when talking about the separability of CP components is whether the bare nominal element can be separated from the verb by real syntactic material, and, if yes, what are the parameters that favor this possibility.

Another issue is the possibility for the nominal element of the CP to be able to project a DP/NP, regardless of its being adjacent to the verb. Recall that according to Karimi-Doostan, only predicative nouns display this property. Especially, concrete nouns like *guš* 'ear' are claimed to always occur in their bare form when part of a CP.

[4]Samvelian (2012) provides numerous similar examples attested in contemporary Persian literature and websites. For more attested examples see also the PersPred Database http://www.perspred.cnrs.fr.

Here again, several counterexamples can be found in the literature, where a concrete noun participating in a CP is nevertheless determined, quantified or modified:

(9) tā čāy xonak šav-ad u sar=i be mahhal=e serqat
until tea cool become.SBJ-3SG he head=INDF to place=EZ burglary
zad
hit.PST.3S
'Until his tea cools, he went to visit the place the burglary had taken place.'[5] (Samvelian 2012: p. 85, ex. 68)

In this attested example from a contemporary Persian novel, the nominal element of the CP *sar zadan* 'to visit' (lit. 'head hit') projects a DP/NP *sar=i* 'a head', since it is determined by the indefinite determiner *=i*. This example and many others mentioned in Samvelian (2012) show that not all concrete nouns are unable of projecting a DP/NP in the context of a CP. The question, as for the previous case, is whether the possibility for a noun to project a DP/NP in the context of a given CP can be correlated to some of its properties.

In this paper, we will focus on the first issue, that is, the separability of the nominal element of the CP. Since the nominal element of the CP is to some extent comparable to a bare direct object, we will compare the possibility for these two elements to be non-adjacent to the verb. Our purpose is to check to what extent the constraint or the preference for the bare nominal element to be adjacent to the verb parallels the tendency for bare DOs to precede the verb immediately. To put it differently, up to now, the issue of separability of the components of a CP has generally been investigated without considering the wider issue of ordering preferences in Persian, especially those involving direct and indirect objects. This is surprising since the literature on Differential Object Marking (DOM) in Persian has extensively discussed the tendency for bare direct objects to be adjacent to the verb, contrary to marked objects, which undergo scrambling.

In the next section, we will present basic word order properties of sentences involving a direct and an indirect object in Persian, with a special focus on recent findings of a series of corpus and experiment-based studies (Faghiri 2016; Faghiri & Samvelian 2014; Faghiri et al. 2014; 2018).

[5] Attested example from *Zan-e ziādi* by J. Al Ahmad (short stories).

4 Bare objects and their position in Persian

The unmarked (neutral or canonical) word order in Persian is SOV. In ditransitive constructions, the ordering of the direct and the indirect object has been claimed to be dependent on the markedness of the direct object: unmarked DOs follow the IO and occur adjacent to the verb, (10a), while marked DOs precede the IO, (10b), and consequently, are separated from the verb (Browning & Karimi 1994; Mahootian 1997; Karimi 2003: among many others). Persian displays DOM. As illustrated in (10b), definite and/or specific DOs are marked by the enclitic *=rā*, which attaches to the last word of the DO. Note also that in formal Persian, there is no overt marker for definiteness, as shown by the fact that *gol* 'flower' has the same form in (10a) and (10b), albeit two different readings with respect to determination.[6]

It should also be noted that in Persian, bare nouns,[7] that is, nouns without any determination or quantification like *gol* in (10a), are not specified for number and therefore can yield a mass reading. Bare objects have either an existential, as in (10a), or a kind-level/generic reading, as in (11).

(10) a. maryam be sārā gol dād
Maryam to Sarah flower give.PST.3SG
'Maryam gave a flower/flowers to Sarah.'

b. maryam gol=rā be sārā dād
Maryam flower=RA to Sarah give.PST.3SG
'Maryam gave the flower to Sarah.'

(11) maryam gol dust dār-ad
Maryam flower friend have.PRS.3SG
'Maryam likes flowers.'

Indefiniteness, on the other hand, is overtly marked in Persian. It can be realized by the enclitic *=i*, as in (12a), by the cardinal *ye*(*k*), as in (12b), or by the

[6]For DOs, the ambiguity is resolved due to the presence of *=rā*. Bare subjects, on the contrary, are ambiguous between an existential or a kind-level generic reading and a definite/specific reading. Thus, in a sentence like *gol ru-ye miz bud*, two readings are available for *gol*: 'A flower/flowers were on the table' or 'The flower was on the table'.

[7]Note that we use the label 'bare' here to refer to nouns that not only appear in their bare form, but also have a non-determined and non-quantified reading. This means that in (10b), *gol* 'flower' is not considered as a bare noun since it receives a definite reading.

combination of these two. Indefinite NPs can have either a specific or a nonspecific existential reading. In the latter case, they are generally *rā*-marked. Contrary to bare nouns, they are always specified for number.

(12) a. maryam gol=i be sārā dād
Maryam flower=INDF to Sarah give.PST.3SG
'Maryam gave a flower to Sarah.'

b. maryam yek gol be sārā dād
Maryam one flower to Sarah give.PST.3SG
'Maryam gave a flower to Sarah.'

More recently, a series of corpus-based and experimental studies (Faghiri 2016; Faghiri & Samvelian 2014; Faghiri et al. 2014; 2018) have allowed for more fine-grained and accurate generalizations on the ordering of complements, which partly go against the previous dichotomous view. In a nutshell, these studies show that the relative order between the DO and the IO: 1) depends on a set of cross-linguistically valid (functional) factors such as degree of determination (or definiteness), phrasal length and animacy; and 2) displays much more variation than previously assumed, implying that it is not empirically justified to posit a canonical order, similar to SOV, for ditransitive sentences. The main conclusions of these studies are:

a. As unanimously claimed in the literature, *rā*-marked DOs strongly prefer to precede the IO, that is, the DO-IO-V word order, and are thus separated from the verb.

b. Bare DOs, by contrast, display a strong preference for the IO-DO-V word order, that is, they follow the IO and appear adjacent to the verb. Importantly, bare modified DOs display more variation and show a relatively less strong preference for being adjacent to the verb.

c. Indefinite (non-*rā*-marked) DOs, however, contrary to what is generally claimed in the literature, are more likely to appear in the DO-IO-V order, that is, they tend to precede the IO. This means that indefinite DOs group with *rā*-marked DOs with respect to their word order preferences rather than with bare objects. Nevertheless, they display more variation and show a relatively less strong preference for the DO-IO-V order.

To sum up, according to these studies, the primary factor that determines the relative position of the DO with respect to IO is the degree of determination (*i.e.*

zero, indefinite, =*rā*-marked or definite) as a cline. This view can capture the fact that DOs located in the middle of the continuum (*i.e.* bare-modified and indefinite DOs) show more ordering variability than the ones located on the two extremities, that is, bare DOs and definite DOs. In other words, the more determined the DO, the more it is likely to be separated from the verb, see Figure 1, adopted from Faghiri (2016: 196).

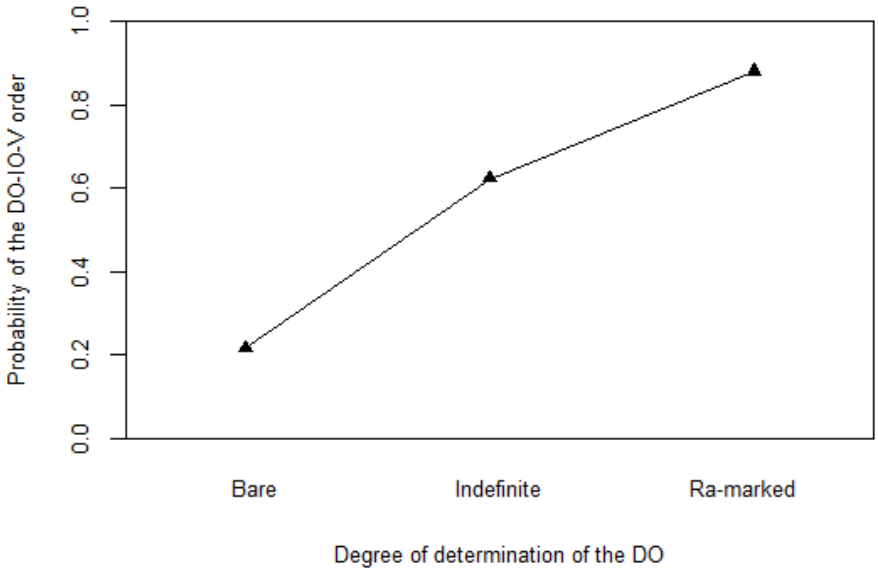

Figure 1: Prediction of default word order on the basis of degree of determination of the DO

Other important findings of these studies are:

d. Phrasal length (or heaviness) also plays a role in ordering preferences. The 'long-before-short' preference is also observed in the preverbal domain in Persian, like in some other SOV languages such as Japanese (Hawkins 1994; Yamashita & Chang 2001). Namely, "heavy" bare DOs, that is, bare-modified DOs, are less likely to appear adjacent to the verb than their 'light' (single word) counterparts.

e. The humanness of the IO favors the IO-DO-V order, which is in line with the general "animate-before-inanimate" preference (**hoberg81a**; Bresnan et al. 2007; Branigan & Feleki 1999; Collins 1995; Kempen & Harbusch 2004; Rosenbach 2002).

5 Empirical study

The review of the literature and the data discussed in previous sections show that in order to get an appropriate account of the (in)separability of CP components we first need to get the empirical facts right. Most of the data provided in theoretical studies rely on "informal" anecdotal grammaticality judgements elicitated without taking necessary methodological precautions and without any control for conflating factors. This undermines the empirical generalizations outlined in these studies, as shown by the abundance of counterexamples, some of which were given previously.

Our aim is to achieve a better understanding of the issue at stake by adopting a quantitative approach that provides us with more reliable data and enables us to investigate and identify different factors that favor (non-)adjacency. The question under study is to what extent the nominal element of a CP, which is formally and syntactically comparable to the direct object (DO) of a lexical verb, is separable from the verb by a prepositional phrase, comparable to the indirect object (IO) of the same verb.

In this section, we present the results of two acceptability judgement experiments carried out via on-line questionnaires and filled out on a voluntary basis by native speakers of Persian living in Iran.

In both experiments, to obtain comparable data on word order variations in the preverbal domain, the questionnaire includes (among other fillers) two additional series of experimental items, besides those for Noun-Verb CPs. One series focuses on the relative order between the (bare) DO and the IO in ditransitive sentences and the other on the relative order of the subject and the (*rā*-marked) DO in transitive sentences.[8] Given that our first experiment serves as a pilot and that our two experiments are similar in many respects, we present and discuss these two additional series of items for the second experiment only.

For Noun-Verb CPs, we compare sentences in which CP components appear in adjacent vs. shifted orders, and manipulate the realization of the nominal ele-

[8]For each participant, the items are ordered in such a way that experimental items of each series are separated by other fillers: items of these different experiments are never presented in a successive order.

ment, comparing bare nouns with indefinite *i*-marked NPs.

We have included a selection of CPs formed by concrete and predicative nouns[9] that take a prepositional argument:[10]

- CPs with concrete nouns: *āb dādan* 'to water' (water give), *āhār zadan* 'to starch' (starch hit), *qazā dādan* 'to feed' (food give), *rang zadan* 'to paint' (paint hit), *rowqan zadan* 'to oil' (oil hit), *vāks zadan* 'to polish' (polish hit), *vāksan zadan* 'to vaccinate' (vaccination hit), *namak zadan* 'to salt' (salt hit).
- CPs with predicative nouns: *fohš dādan* 'to insult' (insult hit), *labxand zadan* 'to smile' (smile hit), *lagad zadan* 'to kick' (kick hit), *ešāre kardan* 'to point' (point do), *kešide zadan* 'to slap' (slap hit), *kalak zadan* 'to trick' (trick hit), *češmqorre raftan* 'to glare' (glare go), *pok zadan* 'to puff' (whiff hit).

All of these CPs display the syntactic pattern given in the canonical order in (13) and illustrated by (14).[11]

(13) N0(=Subj) Prep N1(=IO) N2(=DO) Verb

(14) ali be maryam labxand zad
Ali to Maryam smile hit.PST.3SG
'Ali smiled at Maryam.'

Recall that while we agree with Karimi-Doostan's judgements (see (7c) above) on the impossibility for *guš* 'ear' to project a DP/NP when part of the CP *guš dādan/kardan* 'to listen', we do not endorse his generalization to the whole class of concrete (non-predicative) nouns. There are indeed examples of concrete nouns that can develop into a DP/NP in the context of a CP, such as those included in our selection. *Vāks* 'polish', for instance, in the context of *vāks zadan* 'to polish' (lit. polish hit) (15a), can head a DP/NP and be separated from the verb by a PP (15b)?

(15) a. ali be kafš-hā vāks zad
ali to shoe-PL polish hit.PST.3SG
'Ali polished the shoes.'

[9]Note that our study does not include verbal nouns since, due to their limited nominal properties, they cannot develop into a DP/NP. However, their separability when they form a CP needs to be investigated in forthcoming studies.

[10]The above list includes all CPs used in our second experiment.

[11]We have selected our CPs using the PersPred database (Samvelian & Faghiri 2013).

b. ali behtarin vāks=rā be kafš-hā zad
Ali best polish=RA to shoe-PL hit.PST.3SG
'Ali polished the shoes with the best polish.'

Moreover, the animacy/humanness of the referent of the prepositional argument has been included in our experiments as a control variable, so that we could check whether the humanness of the IO favors the IO-DO-V order, as it is suggested to be the case in ordinary ditransitive constructions in Persian (see page 12).

Our hypothesis is that the CPs of our sample do not differ from ordinary complement-verb combinations concerning word order variations. Therefore, based on the conclusions of Faghiri (2016) presented in Section 4, we predict that:

1. When the nominal element of the CP is realized as an indefinite NP, semantic relatedness favors the adjacent order, while the NP shift is licensed by the general tendency of indefinite DOs to precede the PP argument.
2. For bare nouns, both factors favor the adjacent order.
3. The phrasal length of the nominal element, that is, adding modification to the noun, favors separation.
4. The humanness of the PP argument favors the adjacent order.

5.1 Experiment 1 (pilot)

5.1.1 Method

In our first (exploratory) experiment, we manipulated the nominal element on three levels: **(a)** bare noun, **(b)** indefinite *i*-marked and **(c)** modified indefinite *i*-marked. We prepared our material in a way to have a relatively natural and acceptable sentence with all three forms of the nominal element in the condition of adjacent orders. To this end, we added a continuation to our target sentence, as in (16), in particular, to improve the acceptability of sentences with indefinite *i*-marked nominal elements.

We prepared 24 experimental items in six conditions according to Table 1. In half of our stimuli, the PP argument is animate, as in (16) and (17), and in the other half, it is inanimate, as in (18). In 6 items, the nominal element is a concrete noun.The PP argument is animate only in one, *qazā dādan*, ex. (17). For the sake of space, only one version of each example is given here, the version corresponding

to condition 6 in Table 1, on the basis of which other versions can be constructed straightforwardly.

Table 1: Experiment 1: Conditions

			Order (adjacent vs. shifted)	
Type of the nominal element:			[PP][NP]	[NP][PP]
bare	ex. *fohš*	'insult'	1	4
i-marked	ex. *fohš=i*	'an insult'	2	5
modified *i*-marked	ex. *fohš=e rakik=i*	'a vulgar insult'	3	6

(16) sahar [fohš=e rakik=i] [be sārā] dād va u=rā
Sahar insult=EZ vulgar=INDF to Sarah give.PST.3SG and him=DOM
asabāni kard
angry do.PST.3SG
'Sahar launched a vulgar insult to Sarah and made her angry.'

(17) ali [qazā=ye sabok=i] [be bačče-hā] dād va anhā=rā be pārk
Ali food=EZ light=INDF to child-PL give.PST.3SG and they=RA to park
bord
take.PST.3SG
'Ali gave the children some light food and took them to the park.'

(18) nima [vāks=e siāh=i] [be kafš-hā] zad va anhā=rā
Nima polish=EZ to shoe-PL black=INDF hit.PST.3SG and they=RA
pušid
wear.PST.3SG
'Nima applied some some black polish on the shoes and put them on.'

The experiment was carried out via a web-based questionnaire (on *Ibex Farm* (Drummond 2013)) filled out by 37 native speakers. Participants were asked to rate each sentence on a Likert scale from 1 (absolutely unacceptable) to 7 (completely acceptable).

5.1.2 Results

Figure 4 provides the box and whisker diagram of the distribution of rate by order for the three realizations of the nominal element.

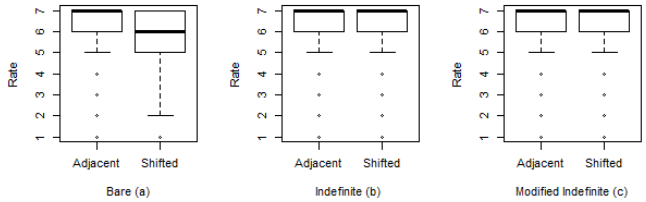

Figure 2: Experiment 1: Distribution of rate by two experimental factors

The statistical analysis of the results showed a significant difference in the rates between adjacent (mean=6.32, SD=1.36) and shifted orders (mean=5.47, SD=1.71) only for bare nouns; $t(36)=5.05$, $p<0.001$. The effect is however of medium size (Cohen's $d=0.53$) and shifted orders are overall rated as acceptable, as we see in Figure 4. For *i*-marked (modified) NPs, both orders are similarly rated as highly acceptable, with mean rates above 6 in all conditions, and we did not find any effect of the phrasal length.

Concrete nouns of our sample display similar rating distributions. However, we will analyze this factor more thoroughly in the second experience, in which the number of items is balanced for concrete and predicative nouns.

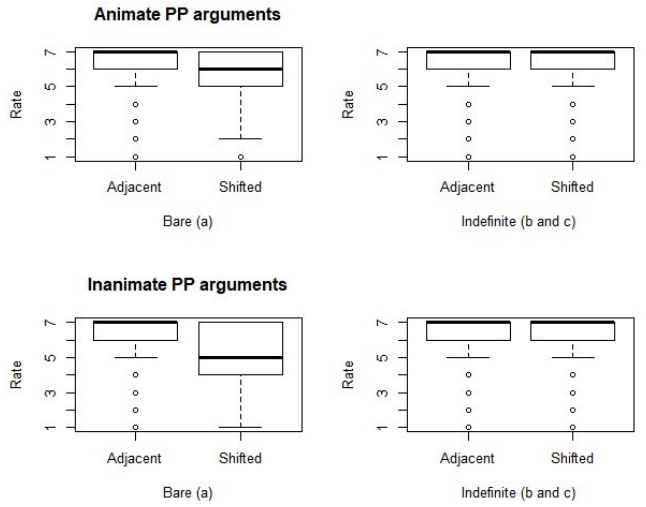

Figure 3: Experiment 1: Distribution of rate by animacy (of the PP)

Interestingly, the humanness of the PP shows an impact on the ratings of sen-

tences with bare nouns. As we can see in Figure 3, animate PPs disfavor the shift more than inanimate PPs do. The statistical analysis, using a linear mixed-effects regression model with order and animacy as fixed effects and items and participants as random effects, showed a small but significant interaction between the two factors (Est.= 0.26, SE=0.07, t=3.44, p<0.01).

Overall, these results are in line with our predictions. However, ratings of 'non-canonical' sentences are surprisingly high, that is, rates below 4 are infrequent. The fact that these sentences were not rated as unacceptable can follow from Faghiri (2016) and Faghiri et al.'s (2018) observations that the relative order between the NP and PP arguments is a matter of soft constraints rather than a syntactic (phrase structure) rule. Hence, while there is a clear bias in production towards a given order, speakers do not consider the alternative order as unacceptable (or ungrammatical), and even, in some cases, consider them as equally acceptable. Nevertheless, to make sure that these results are not due to an experimental confound, we replicated this experiment with a more careful protocol.

5.2 Experiment 2

5.2.1 Method

In this experiment, we chose to keep lexical differences between items to a minimum level:

1. Given that in Experiment 1 we did not find any differences between modified and single-word *i*-marked nominal elements, we removed the modified *i*-marked condition and manipulated the nominal element on two levels, bare vs. indefinite *i*-marked.

2. Contrary to the previous experiment, we kept the sentence simple, that is, without any continuation.

We prepared 16 experimental items (15 from the previous experiment) in four conditions (see Table 2), as illustrated examples (19) – (22). In half of the stimuli, CPs are built with concrete nominal elements, and in the other half, with predicative nouns (see the list on page 14). Two items with concrete nouns are built with animate PP arguments, ex. (19), and six with inanimate PP arguments, ex. (20). Two items with predicative nouns are built with inanimate PP arguments, ex. (22), and six with animate PP arguments, ex. (21). For the sake of space, only one version of each example is given here, the version corresponding to condition 4 in Table 1, on the basis of which other versions can be constructed straightforwardly

Table 2: Experiment 2: Conditions

			Order (adjacent vs. shifted)	
Type of the nominal element:			[PP][NP]	[NP][PP]
bare	ex. *qazā*	'food'	1	3
i-marked	ex. *qazā=i*	'some food'	2	4

(19) ali [qazā=i] [be bačče-hā] dād
Ali food=INDF to child-PL give.PST.3SG
'Ali gave some food to the children.'

(20) maryam [āb=i] [be bāqče] dād
Maryam water=INDF to garden give.PST.3SG
'Maryam (lit.) gave some water to the garden.'

(21) sārā [labxand=i] [be mehmān-hā] zad
Sarah laughter=INDF to guest-PL hit.PST.3SG
'Sarah (lit.) gave a smile to the guests.'

(22) omid [lagad=i] [be dar] zad
Omid kick=INDF to door hit.PST.3SG
'Omid gave a kick to the door.'

Beside these target sentences, our stimuli included four series of control items as fillers (two series of unacceptable control sentences and two series of experimental items on word order variation):

1. 8 sentences with clear grammaticality violations, such as (23).

   (23) * … dišab bārān=rā ziād āmad
   last-night rain=RA very come.PST.3SG
   Intended:'… it rained a lot last night.'

2. 2 sentences similar to the example (7c) above by Karimi-Doostan:

   (24) … amir [guš=e bā-deqqat=i] [be mo'alem] dād
   Amir ear=EZ careful=INDF to teacher give.PST.3SG
   Intended:'… Amir listened carefully to the teacher.'

(25) … neda [češm=e moztarebi=i] [be čamedān] andāxt
Neda eye=EZ worried=INDF to suitcase launch.PST.3SG
Intended:'… Neda looked worriedly at the suitcase.'

3. 8 experimental items, similar to (26), focusing on the relative order between the subject and the *rā*-marked DO in prototypical transitive sentences.[12]

(26) a. … omid maryam=rā nārāhat kard
… Omid Maryam=RA hurt do.PST.3SG
'… Omid hurt Maryam.'
b. … maryam=rā omid nārāhat kard

4. 16 experimental items, similar to (27) and (28), focusing on the relative order between the IO and a bare DO[13] with control for the humanness of the IO.[14]

(27) a. … sar=e miz gol be-gozār-and
on=EZ table flower SBJ-put.PRS-3PL
'… (they) put flowers on the table.'
b. … gol sar=e miz be-gozār-and

(28) a. … barā=ye soxanrān čāy bi-āvar-and
for=EZ speaker tee SBJ-bring.PRS-3PL
'… (they) bring tee for the speaker.'
b. … čāy barā=ye soxanrān bi-āvar-and

The remaining 30 fillers covered a range from highly acceptable to less acceptable sentences.

The experiment was carried out via the same web-based questionnaire. However, unlike the previous experiment, we opted for an 11-point scale from 0 (absolutely unacceptable) to 10 (completely acceptable), which we consider to be

[12] These items are taken from Faghiri's (2016) sentence completion experiment on transitive sentences (see Experiments T1 (pp. 197–204)).

[13] In this experiment, we also also manipulated the phrasal length of the DO, comparing bare and bare modified nouns. Here we only discuss the data for bare DOs.

[14] These items are taken from Faghiri's 2016 sentence completion experiments on ditransitive sentences (see Experiment D2 (pp. 188–193) and Experiment D4 (pp. 188–193)).

more natural for our participants than a 7-point scale. Also, the rating task was followed by comprehension questions in 40 filler items.

116 monolingual speakers of Persian living in Iran filled out the questionnaire. We discarded answers from three participants who had more than 10% of wrong answers to comprehension questions and/or rated clearly ungrammatical sentences as acceptable. Our final dataset hence contained a total number of 1808 observations.

5.2.2 Results

The box and whisker diagram of the distribution of rate by our two experimental factors for our target items is given in Figure 5 and for our unacceptable control items (1 and 2 on page 19) in Figure 5.

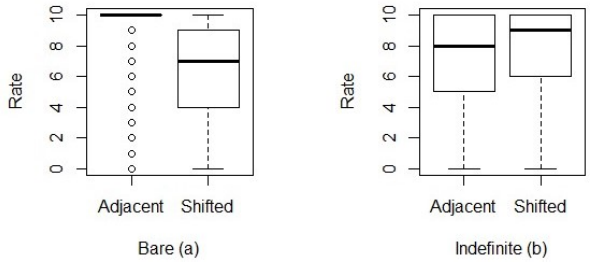

Figure 4: Experiment 2: Distribution of rate by two experimental factors

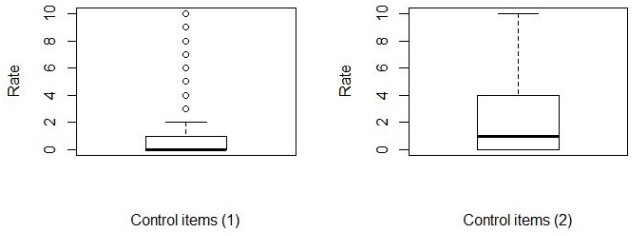

Figure 5: Experiment 2: Distribution of rate for clearly unacceptable control items

We can see that the distribution of rate in our new data is not substantially dif-

ferent from Experiment 1. Here again, the distribution of rates is (almost)[15] identical for both orders in the case of indefinite *i*-marked NPs and the analysis of the results shows a significant decrease for shifted orders (mean=6.49, SD=3.07) compared to adjacent orders (mean=9.27, SD=1.87) only for bare nouns; t(112)=13.43, p<0.001. The effect size is large (Cohen's d=0.96) and much more important than what we had previously. Nevertheless, sentences in the shifted order are still not rated as unacceptable; the median rate is 7. Compare the distribution of rate in target items with our clearly unacceptable controls where both mean and median are very low: respectively 2.4 and 1 for first control items, and 1.2 and 0 for the second ones. It is also insightful to take a closer look into the frequency distribution of rate for adjacent vs. shifted orders, compared to our unacceptable controls (see Figure 6). In sharp contrast to the latter, high scores remain the most frequent rates for shifted orders and the mode is still 10. Indeed, we do not have a bi-modal distribution with some speakers rating these sentences as totally unacceptable and others as perfectly acceptable. Speakers mostly tend to rate these sentences as equally acceptable or slightly less acceptable than canonical sentences.

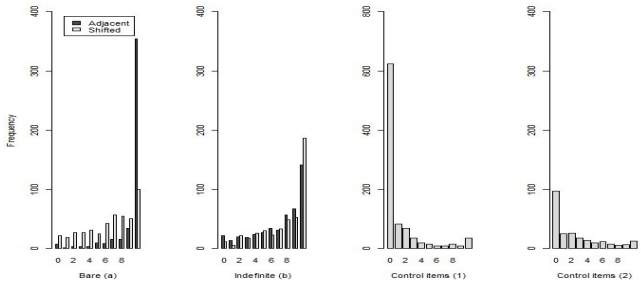

Figure 6: Experiment 2: Frequency distribution of rate for target items vs. unacceptable control items

At this point, let us compare these data with our two other series of experimental items on word order variations, that is, 1) the relative order between the (bare) DO and the IO, and 2) the relative order between the Subject and the DO in prototypical transitive sentences (see box and whisker diagrams in Figure 7). In both cases, we find a significant decrease in the mean rate for "non-canonical" orders as well. However, the effect sizes are smaller and "non-canonical" orders are rated relatively better than what we observe for CPs (with bare nominal elements). In the case of the relative order in transitive sentences, the effect is of

[15]Interestingly, the mean rate is slightly but significantly better for shifted orders: 7.59 (SD=2.85) vs. 7.14 (SD=3.05); t(112)=3.24, p<0.01.

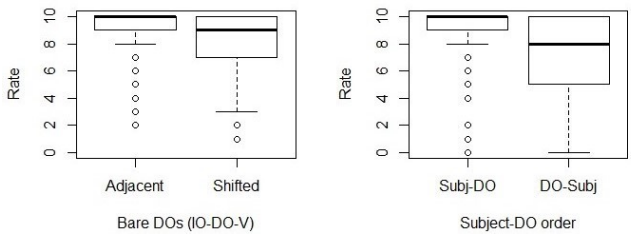

Figure 7: Experiment 2: Distribution of rate for word order variation control items

medium size (Cohen's d=0.73), the difference between the mean rate for canonical (Subj-DO) and non-canonical (DO-Subj) orders is less than 2 points (9.19 vs. 7.34; t(112)= 10.46, p<0.001), and the median rate for non-canonical orders is 8. Interestingly, for bare DOs, the effect size is small – half the size we had for bare nouns forming a CP (Cohen's d=0.48). The difference between the mean rate for adjacent and shifted orders is almost 1 point (9.33 vs. 8.42; t(112)= 7.48, p<0.001) and the median rate for shifted orders is 9.

Finally, let us consider the effect of our two control factors: 1) the type of the nominal element and 2) the humanness of the PP argument. Figures 8 and 9 provide the same box and whisker diagrams of the distribution of rate, respectively, for concrete vs. predicative nominal elements, and for animate vs. inanimate PP arguments.

Recall, however, that these two factors are correlated in our design. Hence, we need to look at the linear mixed-effects model (LMM) analyses of the data (Baayen et al. 2008) in order to be able to capture the effect of these two factors on acceptability judgements independently and in interaction with order. To this end, rates were entered into a mixed-effect linear regression model using the lme4 package (Bates et al. 2012) of the R statistics software (R Core Team 2016). We ran two separate models, one including only bare nouns and the other indefinite *i*-marked NPs. In each model, the experimental factors are included as fixed effects, with sum-coded contrasts.[16] We fitted the full variance-covariance structure of random effects for both items and participants, justified by the design. Table 3 presents the summaries of both models for fixed effects. The results are as following:

1. The estimated mean (baseline) rate across all factors is above 7 in both

[16]Order: Adjacent=1, Shifted=-1; Animacy: Animate=1, Inanimate=-1: Noun-Type: Predicative=1, Concrete=-1.

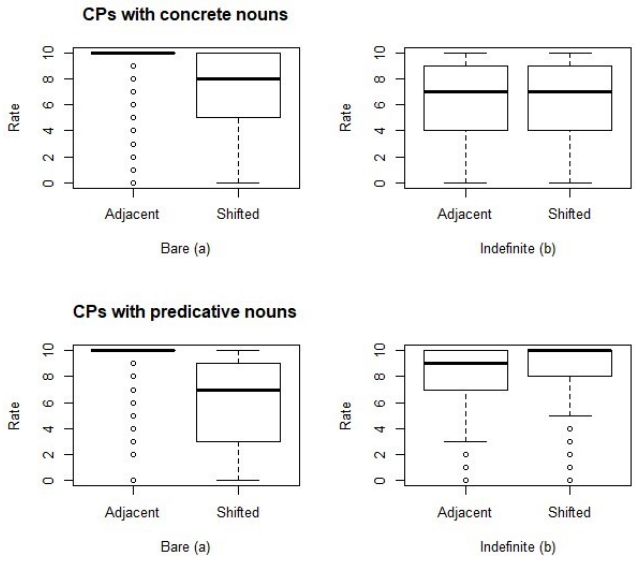

Figure 8: Experiment 2: Distribution of rate by (semantic) type of the nominal element

cases (7.81 and 7.29, for bare nouns and indefinite NPs respectively).

2. As expected, there is a significant and relatively important main effect of order for bare nouns: the difference in the (estimated) mean rates between adjacent and shifted orders is about 3 points. However, in the case of indefinite NPs, the effect of order while significant is very small and, interestingly, works in the opposite direction. The difference in the (estimated) mean rates between shifted and adjacent orders is only about 0.5 point.

3. There is a significant but rather small interaction between order and animacy for both bare nouns and indefinite NPs: with shifted orders, inanimate PPs yield slightly better rates than animate PPs.

4. There is no interaction between Order and Noun-type, neither for bare nouns nor for *i*-marked NPs. For the later, however, Noun-type has a significant and relatively important main effect on rate: predicative nominal elements are rated better than concrete ones regardless of order. The difference in the (estimated) mean rates between the two noun types is about 2 points.

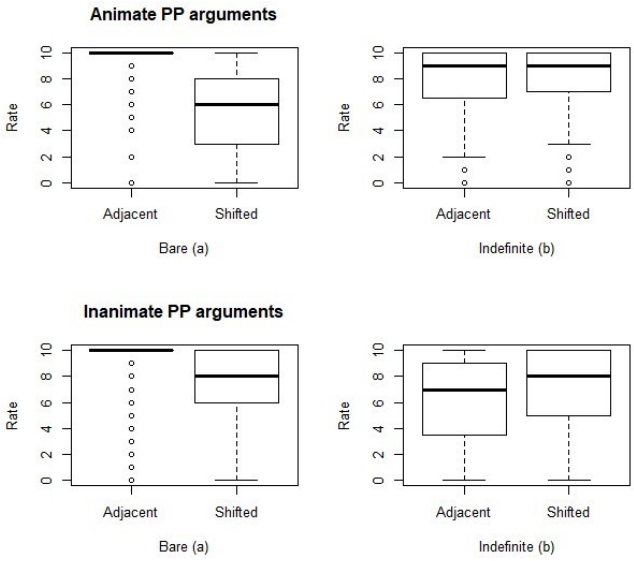

Figure 9: Experiment 2: Distribution of rate by animacy (of the PP argument)

5.3 Main findings

The main findings of our experimental study are:

1. Sentences in which bare nouns forming a CP appear separated form the verb by the PP argument are not considered to be ungrammatical by native speakers, but only less acceptable than sentences in which they appear adjacent to each other. However, in comparison, ordinary ditransitive sentences in which the bare noun is separated from the verb by the PP argument are rated better.

2. When the nominal element of a CP is realized as an indefinite *i*-marked NP, sentences in which the nominal element is separated from the verb by the PP argument are considered slightly more acceptable.

3. The predicative nature of the noun forming a CP has no effect on ordering preferences. In other words, speakers accept sentences in which concrete nouns are separated from the verb in the same manner as they accept those with predicative nouns. In the meanwhile, as expected the ability of the nominal element of a CP to develop a DP projection is affected by its

Table 3: Experiment 2: Results of LMM analyses

Bare nouns (a)

	Estimate	Std. Error	df	t-value	p-value	
Intercept	7.81	0.24	26.03	23.74	<0.001	***
Order [Adjacent=1]	1.47	0.11	40.12	12.64	<0.001	***
Animacy [Animate=1]	-0.21	0.24	14.25	-0.91	0.38	
NounType [Predicative=1]	-0.10	0.23	13.48	-0.45	0.66	
Order:Animacy	0.40	0.10	14.73	3.85	<0.01	**
Order:NounType	0.02	0.10	11.54	0.17	0.87	

Indefinite *i*-marked NPs (b)

	Estimate	Std. Error	df	t-value	p-value	
Intercept	7.29	0.26	46.16	28.16	<0.001	***
Order [Adjacent=1]	-0.23	0.09	11.65	-2.61	<0.05	*
Animacy [Animate=1]	0.04	0.22	14.79	0.21	0.84	
NounType [Predicative=1]	1.07	0.22	16.97	4.78	<0.001	***
Order:Animacy	0.35	0.13	18.13	2.80	<0.05	*
Order:NounType	-0.07	0.11	-0.59	0.17	0.57	

predicative nature: when the nominal element was *i*-marked, CPs of our sample formed by predicative nouns were rated better than those formed by concrete nouns.

4. The humanness of the intervening PP argument disfavors the separability of CP components: sentences in which the nominal element precedes the PP were rated slightly better when the PP argument was inanimate than when the PP argument was human.

In a nutshell, the findings of our study contradict all previous claims on the inseparability of CP components (Goldberg 1996; Karimi-Doostan 1997; 2011: among others) and suggest not only that "real" syntactic material can interrupt a noun-verb CP but also that ordering preferences in CPs are comparable to those observed in ordinary complement-verb combinations, semantic relatedness and collocationality put aside.

Before closing this section, it is important to discuss a previous quantitative evaluation of Karimi-Doostan's claim on the (in)separability of CP components, which arrives at a different conclusion, partially at odds with the conclusions of our study.

In a recent paper on language processing, Safavi et al. (2016) use separable Persian CPs to test the predictions of different accounts of locality effects and follow Karimi-Doostan's classification to select separable CPs. In order to make sure that the CPs included in their experimental material are separable for native speakers they carried out a preliminary norming acceptability rating experiment to test the relative acceptability of "separable" vs. "inseparable" CPs (2016: 4). They had 50 native speakers rate three sets of 36 sentences with CPs from each class, following a between-items design with three conditions: (a) verbal nouns, (b) predicative nouns and (c) non-predicative nouns, and report the following mean rates on a 7-point Likret scale, respectively: 3.23 (Q1=1, Q3=5), 6.08 (Q1=6, Q3=7), and 3.12 (Q1=1, Q3=5), that suggest a clearcut distinction between "separable" and "inseparable" CPs in support of Karimi-Doostan's classification.

Nevertheless, a close look into their stimuli, provided on-line as supplementary material,[17] shows that they did not perform a systematic (minimally paired) comparison across the three conditions. An example of items in each condition is given in (29).[18]

(29) a. hamkār=am nāme=i *ersāl* [be man] kard
colleague=CL.3SG letter=INDF sending to me do.PAST.3SG
'My colleague sent me a letter.'

b. maryam *arezu=i* [barā=ye man] kard
Maryam wish=INDF for I do.PAST.3SG
'Maryam made a wish for me.'

c. golnaz *otu* [be lebās=aš] zad
Golnaz iron be dress=CL.3SG hit.PAST.3SG
'Golnaz ironed her dress.'

We notice that while the indefinite *i*-marked form is used in sentences with predicative nouns, as in (29b), the bare form is used for the other classes, as in (29a) and (29c). This is not surprising given that, as we have seen in Section 3, the issue of (in)separability is entangled with the realization of the NP (or the ability of the nouns to develop a DP/NP projection) in Karimi-Doostan's view. However, this design makes the comparison between the three conditions meaningless.

Putting aside verbal nouns that cannot develop a DP/NP projection, we have seen that in a number of CPs involving a non-predicative noun, the nominal element can develop a DP/NP projection and be separated from the verb. However,

[17] Accessible via the following link: http://journal.frontiersin.org/article/10.3389/fpsyg2016.00403
[18] Glossing and translations are ours.

Safavi et al. did not control for this property in their design and in a number of their items involving a non-predicative noun, such as *otu zadan* 'to iron' (iron hit) in (29c), the nominal element can appear in the *i*-marked form, as illustrated in (30), while they include also CPs such as *guš dādan* 'to listen' (ear give).

(30) golnaz *otu=i* be lebās=aš zad
Golnaz iron=INDF to dress=CL.3SG hit.PAST.3SG
'Golnaz ironed her dress.'

Note that the stimuli include only one version of each item: a sentence in which the noun is separated from the verb by a prepositional phrase and there is no control on the function and semantics of the intervening PP. As a consequence, while their data serve the initial purpose of their norming pretest; they do not provide evidence for the inseparability of CPs with non-predicative nouns (in opposition to CPs with predicative nouns).

6 Conclusions

The experimental data presented in the previous section 1) provide additional support, along with the attested counterexamples given in Section 3 and in Samvelian (2012), that the nominal element of a CP, whatever its form and its type, can be separated from the verb by syntactic material, and 2) suggests that the issue of separability in CPs cannot be studied separately from word order preferences involving the verb and its complements in ordinary transitive and ditransitive constructions.

However, our study constitutes a first step in the study of the issue of separability with quantitative and experimental methods. Further studies are needed in order to investigate several points that we did not address in this paper:

Production data. Our study suggests that speakers have an important tolerance for sentences in which the bare nominal element is separated from the verb by the PP argument of the CP, since, as explained in the previous section, acceptability rates stay high, that is, clearly above the baseline. In order to have a more accurate picture, the acceptability judgement data must be completed by production data, including corpus studies.

Separation by other constituents than PP arguments. Our experiments were designed with sentences in which the intervening element was the PP argument of the CP since our purpose was to assess Karimi-Doostan's claim on separability. The possibility for other constituents, such as adverbials, to intervene between

the nominal element and the verb must be investigated in forthcoming studies. However, we should insist on the fact that such an investigation cannot be carried out without considering the same kinds of possibilities in ordinary Object-Verb combinations, particularly in the case of bare objects. Recall from Section 4 that, as mentioned in several studies, a bare object of lexical verbs displays also a limited degree of autonomy with respect to the verb and tend to occur adjacent to the latter.

Separability and DP projection. All examples in our data were designed with nouns that can project NP/DPs, be they predicative or concrete, since the purpose was not only to check the possibility for bare nouns to be separated from the verb but also to study the role of the degree of determination in ordering preferences. However, not all concrete nouns can project a DP/NP when forming a CP. Recall the example of *guš* 'ear' in *guš dādan/kardan* 'to listen' (ear give/do) given by Karimi-Doostan. Although, we have not included these cases in our experiments, it seems safe to consider that their behavior (as bare nouns) should not be different from those that can project a DP/NP in the context of a given CP. Note that examples of separation for *guš dādan/kardan* abound in the literature. Here are a few of them:

(31) abbas nārāhat va pažmorde bud. guš be mādar=aš dād…
Abbas sad and unhappy was ear to mother=CL.3SG give.PST.3SG
'Abbas was sad and unhappy. He listend to his mother…' (Ali Ašraf Darvišian, *Ǯang be revāyat-e bačče hā*, p. 30)

(32) guš be harf=e man bo-kon … va baqi=rā vel-kon…
ear to speech=EZ me IMP-do.2SG and rest=RA leave.IMP.2SG
'Listen to me (…) and let go of the rest…' (Širin Sami'i, *Bibi va touti*, p. 44)

Apart from non-projecting concrete nouns, we also excluded verbal nouns, ex. *ersāl* 'sending' from our study. Recall that the latter display limited nominal properties and can never be determined, in the context of a CP or not. It seems that verbal nouns resist more separation than predicative and concrete nouns. Although this fact needs to be checked by further empirical studies, it would not be surprising a priori. Indeed, the problematic status of these "nouns" can account for the fact that they are not perceived as direct objects and consequently are not subject to the same ordering variations.

To conclude, Persian CPs, like other types of multiword expressions in various languages, illustrate a case of deviation from the one-to-one mapping of form and

meaning. Despite the fact that they realize a single semantic unit, their components nevertheless enjoy the mobility granted to members of "ordinary" verb complement syntactic constructions and undergo the same constraints with respect to the linear order. The semantic bond plays nevertheless a role in granting CPs hallmarks of "wordhood", favoring their adjacency, among other things.

Acknowledgements

References

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* 59(4). 390–412.

Bates, Douglas, Martin Maechler & Ben Bolker. 2012. *Lme4: linear mixed-effects models using s4 classes*.

Branigan, Holly P & Eleonora Feleki. 1999. Conceptual accessibility and serial order in greek language production. In *Proceedings of the 21st Conference of the Cognitive Science Society*, 96–101.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Kramer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Arts & Science.

Browning, Marguerite & Ezat Karimi. 1994. Scrambling to object position in Persian. In Norbert Corver & Henk van Riemsdijk (eds.), *Studies on scrambling: movement and non-movement approaches to free word-order phenomena*, 61–100. Berlin: Mouton de Gruyter.

Collins, Peter. 1995. The indirect object construction in English: an informational approach. *Linguistics* 33(1). 35–50.

Dabir-Moghaddam, Mohammad. 1997. Compound verbs in Persian. *Studies in the Linguistic Sciences* 27(2). 25–59.

Drummond, Alan. 2013. *Ibex farm (free hosting for Internet-based experiments)*. http://www.spellout.net/ibexfarm. Version 0.3.

Faghiri, Pegah. 2016. *La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique*. Université Sorbonne Nouvelle dissertation.

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 & 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.

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2018. Is there a canonical order in Persian ditransitive constructions? Corpus based and experimental studies. In Agnes Korn & Andrej Malchukov (eds.), *Ditransitive constructions in a cross-linguistic perspective*, 165–185. Wiesbaden: Reichert Verlag.

Ghomeshi, Jila. 1996. *Projection and inflection: a study of Persian phrase structure*. University of Toronto dissertation.

Givi Ahmadi, Hassan & Hassan Anvari. 1995. *Dastur zabāne fārsi [Persian grammar]*. Mo'assese farhangi Fātemi.

Goldberg, Adele E. 1996. Words by default: optimizing constraints and the Persian complex predicate. In *Annual proceedings of the Berkeley Linguistic Society 22*, 132–146. Berkeley.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 2001. Why are categories adjacent? English. *Journal of Linguistics* 37(1). 1–34.

Karimi-Doostan, Gholamhossein. 1997. *Light verb constructions in Persian*. University of Essex dissertation.

Karimi-Doostan, Gholamhossein. 2011. Separability of light verb constructions in Persian. *Studia Linguistica* 65(1). 70–95.

Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Oxford: Blackwell Publishing Ltd. DOI:10.1002/9780470758403.ch5

Kempen, Gerard & Karin Harbusch. 2004. A corpus study into word order variation in German subordinate clauses: animacy affects linearization independently of grammatical function assignment. In Thomas Pechmann & Christopher Habel (eds.), *Multidisciplinary approaches to language production*, 173–181. Berlin: Mouton de Gruyter.

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.

Megerdoomian, Karine. 2002. *Beyond words and phrases: a unified theory of predicate composition*. University of Southern California dissertation.

Mohammad, Jan & Simin Karimi. 1992. Light verbs are taking over: complex verbs in Persian. In *Proceedings of the Western Conference on Linguistics (WECOL) 5*, 195–212.

Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46(2). 601–655.

Rosenbach, Anette. 2002. *Genitive variation in English: conceptual factors in synchronic and diachronic studies*. Berlin: Walter de Gruyter.

Safavi, Molood S, Samar Husain & Shravan Vasishth. 2016. Dependency resolution difficulty increases with distance in Persian separable complex predicates: evidence for expectation and memory-based accounts. *Frontiers in psychology* 7. 403. DOI:10.3389/fpsyg.2016.00403

Samvelian, Pollet. 2001. Le statut syntaxique des objets nus en persan. French. *Bulletin de la Société de Linguistique de Paris* 96(1). 349–388.

Samvelian, Pollet. 2012. *Grammaire des prédicats complexes : les constructions nom-verbe*. Paris: Hermès sciences–Lavoisier.

Samvelian, Pollet & Pegah Faghiri. 2013. Introducing PersPred, a syntactic and semantic database for Persian complex predicates. In *Proceedings of the 9th Workshop on Multiword Expressions (MWE 2013), NAACL-HLT*, 11–20. Atlanta: Association for Computational Linguistics.

Samvelian, Pollet & Pegah Faghiri. 2014. Persian complex predicates: How compositional are they? *Semantics-Syntax Interface* 1(1). 43–74.

Samvelian, Pollet & Pegah Faghiri. 2016. Re-thinking compositionality in Persian complex predicats. In *Proceedings of the 39th Berkeley Linguistics Society*. Berkeley: Linguistic Society of America.

Vahedi-Langrudi, Mohammad-Mehdi. 1996. *The syntax, semantics and argument structure of complex predicates in modern Farsi*. University of Ottawa dissertation.

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: CSLI Publications.

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