Accessibility and word order: The case of ditransitive constructions in Persian

Pegah Faghiri
Université Sorbonne Nouvelle

Pollet Samvelian
Université Sorbonne Nouvelle

Barbara Hemforth
CNRS/Laboratoire de Linguistique Formelle (LLF)

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 217–237

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2014. Accessibility and word order: The case of ditransitive constructions in Persian. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 217–237. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.12.

Abstract

In a most recent corpus study on Persian, Faghiri & Samvelian (2014) found a significant effect of relative length in the ordering preferences between the direct and indirect objects in the preverbal domain corresponding to "long-before-short". They furthermore showed that the position of the direct object mainly depends on its degree of determination, and put into question the broadly accepted dual view based solely on differential object marking. In this paper, we provide experimental evidence in support of these corpus findings and further propose a unified account of ordering preferences between the two objects on the basis of conceptual accessibility.

1 Introduction

Preferences in constituent ordering have often been explained by the widely accepted accessibility-based incremental model of sentence production. In this view, the linear order of constituents reflects the order in which they become available for production, as long as grammar rules do not intervene (e.g. Garrett, 1980; Bock & Levelt, 1994; Kempen & Harbusch, 2003). Constituents that become available at an earlier point in time, can occupy an earlier linear position than constituents emerging later. This view is fully compatible with the "short-before-long" principle (e.g. Wasow, 1997; Stallings et al., 1998; Arnold et al., 2000; Wasow, 2002). Short simple constituents can be processed and formulated faster and thus become available for production sooner than long and/or complex ones. However, the predictions of this model have been shown to be incompatible with the long-before-short preference observed in the preverbal domain in head-final languages such as Japanese (Hawkins, 1994; Yamashita & Chang, 2001) and Korean (Choi, 2007)

The mirror-image preference in head-initial and head-final languages was first observed by Hawkins (1994, 2004) who proposed a dependency-based distance-minimizing principle in terms of a theory of parsing efficiency. He proposed the *Early Immediate Constituent* (EIC) principle[1] to account for these seemingly contradictory preferences in head-final and head-initial languages. Yamashita & Chang (2001) provide experimental evidence for "long-before-short" preference in sentence production in Japanese. They further proposed a production-oriented account of these conflicting ordering preferences in the framework of the theory of grammatical coding (Bock & Levelt, 1994; Garrett, 1980).

[†]We would like to thank the audience at the Workshop on Understudied Languages and Syntactic Theory - HPSG21 (University at Buffalo) for their insightful comments, as well as Stefan Müller the editor of this volume. Parts of the present paper have also been presented at the Workshop on Ditransitive Constructions in a Cross-linguistic Perspective, adjacent to the SWL6 Conference at the University of Pavia in September, 2014. We would also like to thank the audience of this venue. 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).

[1]And its more recent version, *Minimize Domains* (MiD).

Their argumentation is mainly based on the fact that in the theory of grammatical encoding decisions about word order depend on the properties of different levels in the production system. Some decisions are more conceptually driven while others depend more on form (Bock, 1982). In the conceptual arena, there is a tendency to place salient elements earlier in sentences while in the form arena, short elements are placed earlier. Long constituents have competing properties. They are semantically richer, due to the extra lexical material which makes them more salient and increases their overall accessibility in the conceptual arena. At the same time, in the form arena, the extra lexical content makes them slower to process and hence less accessible.

The authors suggest that acknowledging language-specific differences in sentence production is the key to a uniform account of word order preferences (also see Chang, 2009). They suppose that the sensitivity of sentence production system to conceptual vs. formal factors can be seen as being language-specific. The production system of Japanese, they argue, is more sensitive to conceptual factors than to form-related ones, contrary to English. This is because Japanese is a far less "rigid" language than English. Japanese has a fairly free word order and allows null pronouns. English, in contrast, has a fairly strict word order that requires all arguments to be overtly present (Yamashita & Chang, 2001, p.54). Moreover, in English Heavy-NP shift happens in the postverbal domain, where it is shown that the verb exerts strong influence, contrary to the preverbal domain (Stallings et al., 1998). These syntactic constraints presumably increase the effect of form-related factors over more conceptual ones. The authors consequently conclude that the Japanese sentence production system, more sensitive to conceptual factors, favors placing long constituents before shorter ones, while in English, more sensitive to form-related factors, placing short constituents before longer ones is favored.

As Japanese, Persian is an SOV language with a fairly free word order and null pronouns. Hence, this language share all properties singled out by Yamashita & Chang (2001) to motivate opposite length-based shifts in Japanese and English. In line with their prediction, in a corpus study on the ordering preferences between the direct (DO) and indirect (IO) objects in the preverbal domain in Persian, Faghiri & Samvelian (2014) have found a significant effect of relative length corresponding to the "long-before-short" preference. Moreover, they have shown that the relative order of the two objects depends mainly on the degree of determination of the DO, which is closely related to discourse status of the latter and hence reflects its conceptual accessibility.

The second aspect of this paper is that it undermines the broadly accepted view of the relative order of the DO and the IO in Persian, see section 3. It is generally assumed that differential object marking determines whether the DO follows or precedes the IO. Yet, the corpus data do not reflect a dichotomous behavior based on differential object marking. The study rather suggests that the position of the DO depends on its degree of determination.

In this paper, we provide experimental evidence in support of these corpus findings and propose a unified account of the relative order between the DO and the

IO in Persian on the basis of conceptual accessibility. The remainder of this paper is organized as follows. In the next section, we present an overview of Persian focusing on properties relevant for this study and in section 3, the prevailing view on the position of direct object. The corpus data of (Faghiri & Samvelian, 2014) will be summarized in section 4. In section 5, we present our experimental study, and in section 6, our unified account.

2 A Brief Overview of Persian

2.1 Word Order

The unmarked (neutral or canonical) word order in Persian is uncontroversially SOV (except for sentential complements which are strictly postverbal). Meanwhile, all phrasal categories (other than the VP), namely, NP, PP, and CP are head-initial, as illustrated in (1). Also, note in the same example that Persian does not require all arguments to be overtly realized.

(1) be ān doxtar=e javān ke diruz did-im goft (ke) emruz
to that girl=EZ[2][3] young that yesterday saw-1PL said.3SG that today
nay-āy-ad
NEG-come-3SG
'S/he said to that young girl we saw yesterday not to come today.'

While SOV is the canonical order, all other variations are possible. Although the written language is conservative with regards to the canonical SOV order, the colloquial register exhibits a fair amount of variation. In their corpus study, Faghiri & Samvelian (2014) focus on verb-final constructions. Given that the experimental data presented in this paper is a follow up on their conclusions, we have also kept the focus of this study on verb-final constructions.

2.2 Persian NPs

The relative order of objects in Persian has generally been linked to the differential object marking (DOM) (see section 2.3), which in turn is related to definiteness and/or specificity. This section provides an overview of Persian NPs in this respect.

In formal Persian there is no overt marker for definiteness; only indefiniteness is marked. Furthermore, Persian has what Corbett (2000) calls a *general number*, expressed by the singular form. This means that in Persian the number is not specified for a bare singular noun. These properties have some bearings on the

[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: DOM = differential object marking; EZ = Ezafe.

[3]The *Ezafe*, realized as an enclitic, links the head noun to its modifiers and to the possessor NP (see Samvelian 2007).

readings of NPs. In the remainder of this section, we will discuss the following NP types: bare[4]and bare-modified nouns, indefinite/quantified NPs, and definite NPs.

2.2.1 Bare and Bare-modified Nouns

Bare nouns are non-specified for number and have a nonspecific reading, which can be generic as well as existential:

(2) gorg yek heyvān=e vahši va darande ast
wolf a animal=EZ wild and predator is
'The wolf is a wild and predator animal.'

(3) Maryam ketāb xarid
Maryam book bought
'Maryam bought a book/some books.'

Bare-modified nouns only differ from bare nouns by the presence of a (restrictive) modifier, as in (4).

(4) Maryam ketāb=e akkāsi xarid
Maryam book=EZ photography bought
'Maryam bought a photography book/some photography books.'

2.2.2 Indefinite NPs

Indefiniteness is overtly marked in Persian. It can be realized by the enclitic *=i*, as in (5-a), by the cardinal *ye(k)*[5] 'one', as in (5-b), or by the combination of these two determiners, as in (5-c).[6] Indefinite NPs can have either a specific or a nonspecific existential reading. As we will see, in the DO position the two readings will be differentiated by DOM. These NPs, contrary to bare nouns, are always specified for number.

Indefinite NPs are also formed by numerals or other indefinite quantifiers, as in (6). In this case, the noun remains in the singular form, even when the NP denotes more than one entity, and it cannot take *=i*.

(5) a. gorg=i zuze mi-kešid
wolf=INDF howl IPFV-pulled

[4]It should be noted that since definiteness is not overtly marked, bare singular nouns, that is, nouns occurring alone in their bare singular form with no (overt) determiner or quantifier, may correspond either to a definite and/or anaphoric NP, as in (i), or to a noun without any determination or quantification. By "bare noun" we only refer to the latter.

(i) xoršid dar āsemān mi-deraxš-ad
sun in sky IPFV-shine-3SG
'The sun shines in the sky.'

[5]Pronounced *ye* in colloquial speech. We will use the formal form throughout this article.

[6]The use of the enclitic alone is restricted to the formal language.

b. yek gorg zuze mi-kešid
a wolf howl IPFV-pulled

c. yek gorg=i zuze mi-kešid
a wolf=INDF howl IPFV-pulled
'A (any/certain) wolf was howling.'

(6) čand(=tā)/se(=tā) gorg zuze mi-kešid-and
few(=CLF)/three(=CLF) wolf howl IPFV-pulled-3PL
'A few/three wolves were howling.'

2.2.3 Definite NPs

Definite NPs can either be formed by different definite determiners, like demonstratives, or by no overt determiner, as in (7).[7] Furthermore, bare plural nouns[8] generally trigger a definite reading, as in (8). Note, however, that the plural marking is not incompatible with the indefinite determination *=i* or *yek*, as in (9) (for a discussion of plural marking and definiteness, see Ghomeshi 2003).

(7) (in) šiše emruz šekast
(this) glass today broke
'This/the glass broke today.'

(8) šiše-hā emruz šekast-and
glass-PL today broke-3PL
'The (*Some) glasses broke today.'

(9) yek ketāb-hā=i heyn=e asbābkeši gom šod-and
a book-PL=INDF during=EZ move lost became-3PL
'Some (of the) books get lost during the move.'

2.3 Differential Object Marking

Persian displays differential object marking (DOM),[9] realized by the enclitic *=rā*.[10] Definite and/or specific direct objects are necessarily *rā*-marked. Consequently, non-*rā*-marked direct objects receive an indefinite nonspecific reading, as in (10). DOM is not incompatible with the indefinite determination, as in (11). An indefinite NP like *ketāb=i* when *rā*-marked will receive a specific reading.

[7] It should be noted that colloquial speech displays a definite suffix, realized as *-(h)e*, which marks a noun as being discourse-given or anaphoric, for example, *gorbe-he* 'the cat'.

[8] Persian disposes of several nominal plural suffixes, among them the suffix *-(h)ā* is universal and can systematically be added to any noun to form a plural (for a review of the nominal plural marking see Lazard et al. 2006 and Faghiri 2010, among others).

[9] This designation coined by Bossong (1985) denotes the property of some languages with overt case-marking of direct objects to mark some objects, but not others, depending on semantic and pragmatic features of the object; see also Aissen (2003).

[10] Realized as *=(r)o* in colloquial speech. We use the formal form throughout this paper for the ease of reading.

(10) Maryam ketāb=rā xarid *vs.* Maryam ketāb xarid
Maryam book=DOM bought Maryam book bought
'Maryam bought the book.' *vs.* 'Maryam bought a book/some books.'

(11) Maryam ketāb=i=rā xarid *vs.* Maryam ketāb=i xarid
Maryam book=INDEF=DOM bought Maryam book=INDEF bought
'Maryam bought a (specific) book.' *vs.* 'Maryam bought a book.'

Nevertheless, *rā*-marking cannot be accounted for on the basis of definiteness and specificity only, as illustrated by (12). Furthermore, the use of the enclitic *=rā* is not limited to DOM. It is also used to mark discourse prominence for other non-subject functions, as in (13). Meanwhile, a more detailed discussion is beyond the scope of the present study (for further discussions see Lazard 1982; Meunier & Samvelian 1997; Dabir-Moghaddam 1992; Roberts et al. 2009, among others).

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

(13) emruz=rā dars mi-xān-am
today=DOM lesson IPFV-read-1SG
'As for today, I (will) study.'

2.4 Complex Predicates

Persian has a limited number of simplex verbs, around 250, half of which are currently used by the speech community. The verbal lexicon mainly consists of syntactic combinations, called "complex predicates", also known as Compound Verbs or Light Verb Constructions, including a verb and a non-verbal element, for example, a noun, as in *bāzi kardan* 'to play' (lit. 'play do'), an adjective, as in *derāz kešidan* 'to lay down' (lit. 'long pull'), a particle, as in *bar dāštan* 'to take' (lit. 'PARTICLE have'), or a prepositional phrase, as in *az dast dādan* 'to loose' (lit. 'of hand give'). New "verbal concepts" are regularly coined as complex predicates rather than simplex verbs (see Samvelian 2012; Samvelian & Faghiri 2013, 2014, among many others).

Although, Persian complex predicates are multiword expressions and thus display some lexical properties such as lexicalization, they display all properties of syntactic combinations, including some degree of semantic compositionality. Hence, as Samvelian (2001, 2012) extensively argues, it is impossible to establish a clearcut distinction between (prep-)noun-verb complex predicates and "ordinary" object-verb combinations. In other words, the differentiation is better reflected by a continuum from highly lexicalized complex predicates to ordinary complement-verb combinations rather than a categorical distinction. Following this observation and given the impossibility of the task, Faghiri & Samvelian (2014) did not attempt to apply any filter to exclude complex predicates from their dataset. In our experimental study, we only included combinations that could hardly be qualified as complex

predicates and would safely be located on the other extremity of the continuum.

3 The Position of the Direct Object

The relative order of the DO and the IO in Persian is generally assumed to depend on *rā*-marking. It is broadly admitted that in a neutral word order *rā*-marked DOs precede the IO while non-*rā*-marked DOs are adjacent to the verb (Mahootian, 1997; **?**; Roberts et al., 2009, among others). Hereafter, we refer to this hypothesis as the DOM criterion.

Several theoretical studies, mainly in the generative framework, further argue for the existence of two different syntactic positions for the DO depending on its markedness or more precisely its specificity (Ghomeshi, 1997; Karimi, 2003; Ganjavi, 2007, among others). To give an example, (14) illustrates the two positions assumed by Karimi (2003, p.105), one of the most frequently cited paper among the above-mentioned. She, furthermore, assumes that a nonspecific or in other words non-*rā*-marked DO can be separated from the verb, that is, can undergo scrambling, only if it has a contrastive focus. The scrambling of specific objects is less constrained, since they can additionally be topicalized.[11] The examples provided by (Karimi, 2003, pp.91–92) to illustrate these claims are given in (15).

(14) a. $[_{VP}\ DP_{[+Specific]}\ [_{V'}\ PP\ V]]$
b. $[_{VP}\ [_{V'}\ PP\ [_{V'}\ DP_{[-Specific]}\ V]]]$

(15) 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 barā mā ye še'r az Hafez mi-xun-e
Kimea often for us a poem from Hafez IPFV-read-3SG
'It is often the case that Kimea 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
'It is often the case that Kimea 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
'It is often the case that Kimea reads a (particular) poem by Hafez for us.'

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

[11] Karimi (2003, pp.106–111) assumes that discourse functions trigger movement in Persian and the landing site of a scrambled object is the specifier of a functional head, such as Topic or Focus.

Grammarians have also formulated generalizations about the canonical position of the DO, which are mostly in accordance with the DOM criterion. However, some additionally establish a distinction between unmarked DOs, depending upon the presence of the indefinite marked *-i*. Givi Ahmadi & Anvari (1995, p.305), for instance, state that *rā*-marked DOs should precede the IO, non-*rā*marked DOs should follow the IO, and *i*-marked (non *rā*-marked) indefinite DOs can either follow or precede the IO.

Faghiri & Samvelian (2014) have conducted the first corpus-based study to investigate the ordering preferences between the DO and the IO. Their study undermines the DOM criterion. Namely, in their corpus data, indefinite (non-*rā*-marked) DOs are in majority non adjacent to the verb, and hence group with *rā*-marked DOs and not with bare and bare-modified DOs.

The experimental study we present in this paper, see section 5, is a follow up on the corpus findings of Faghiri & Samvelian (2014). In the next section, we briefly present their data and results.

4 Summary of Corpus Results

4.1 The Dataset

The study is conducted on the Bijankhan corpus, a corpus collected from daily news and common texts, in particular, the newspaper *Hamshahri*, of about 2.6 million tokens, manually tagged for part-of-speech information.[12]

To constitute their dataset, the authors have selected the potentially ditransitive verbs of the corpus (122 verb types), corresponding to 42,550 tokens and have identified relevant sentences, that is, sentences matching either of NP PP V or PP NP V patterns (without take into consideration the preceding constituents of the sentence) in two separate samples : 1) a random sample of 2000 tokens out of this subset, and 2) a sample including all occurrences of two typically ditransitive low frequency verbs of the corpus (*rixtan* 'to pour' and *ferestādan* 'to send'; 219 and 254 tokens, respectively), as well as a random sample out of all occurrences of two high frequency typically ditransitive verbs (*gereftan* 'to give' and *dādan* 'to take'; 10494 and 6849 tokens, respectively). This second dataset contains 905 tokens.

4.2 The DOM Criterion

The data is annotated for the DO type according to a fine-grained classification based on the degree of determination of the NP. Marked and bare DOs correspond, respectively, to the highest and the lowest degree of determination for an NP in the DO position in Persian. For intermediate cases, that is, non-bare non-*rā*-marked DOs, the authors have separated determined NPs, that is, quantified or indefinite NPs, from non-determined NPs, that is, bare-modified NPs. Recall that the latter

[12]http://ece.ut.ac.ir/dbrg/bijankhan/

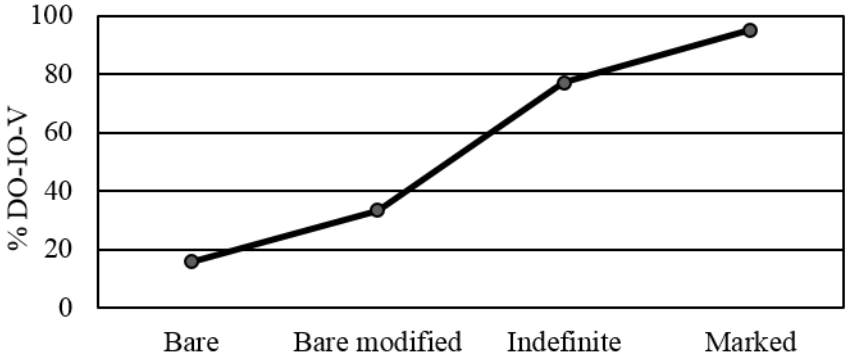

Figure 1: Number of DO-IO-V orders in % by DO-type

only differ from bare nouns by the presence of a modifier. Four DO types are hence defined: Bare, Bare-modified, Indefinite, and Marked.

The ordering preferences observed for marked DOs, on the one hand, and bare and bare-modified DOs, on the other hand, conform to the DOM criterion. Marked DOs prefer the DO-IO-V word order. Bare and bare-modified DOs prefer the IO-DO-V word order. Indefinite DOs, however, contrary to what is expected from the DOM criterion, prefer the DO-IO-V word order. Indeed, being non-*rā*-marked, it is expected for these DOs to group with bare and bare-modified DOs and prefer the IO-DO-V word order. Upon this observation, Faghiri & Samvelian (2014) call into question the DOM criterion and propose an account that subordinates the position of the DO to its degree of determination rather than to its markedness. This account, they claim, has furthermore the benefit of capturing the variation in the strength of the preference.

The degree of determination is, indeed, a continuum, and consequently, this account does not predict ordering preferences on a dichotomous basis as it is the case with the DOM criterion. Faghiri & Samvelian (2014) formulate their account as following: The more a DO is determined, the more it is likely to be placed leftward in the sentence and separated from the verb. Or, vice versa, the less a DO is determined, the more likely it is to be placed adjacent to the verb, see Figure 1. Put this way, it is expectable for DOs located in the middle of the continuum to show more variability than the ones located on the two extremities.

4.3 The Relative Length

Faghiri & Samvelian (2014) have also investigated the question of the relative length via their corpus data. They argue that this factor is not relevant for all DOs. On the one hand, marked DOs prefer the DO-IO order regardless of the relative length, and on the other hand, relative length is meaningless for bare DOs, since these DOs are by definition smaller than the IO in number of words. However,

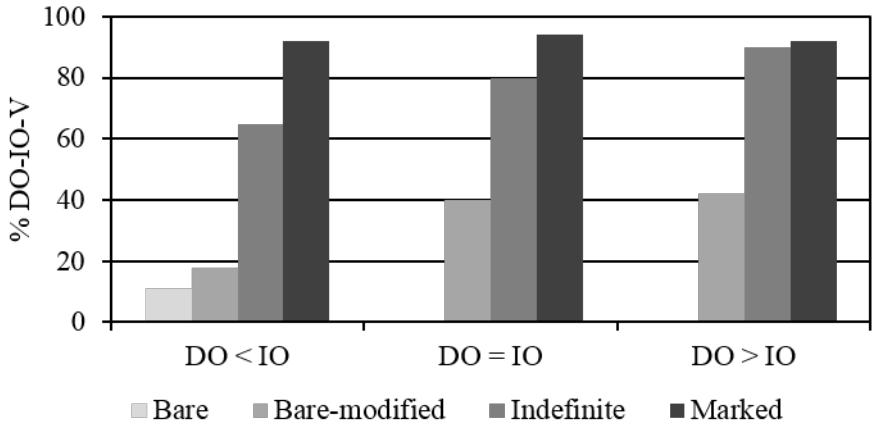

Figure 2: Number of DO-IO-V orders in % by DO-type and Relative Length

the relative length is a relevant factor in the case of indefinite and bare-modified DOs, see Figure 2. Accordingly, the authors have performed mixed-effect logistic regression modeling, on a subset of their dataset excluding marked and bare DOs.

As expected, DO-TYPE has a significant effect ($p < 0.001$). Bare-modified type favors the IO-DO order while Indefinite type favors the inverse, confirming thus that indefinite DOs do not behave like other non-*rā*-marked DOs. In line with our earlier assumptions, REL-LEN also has a significant effect ($p < 0.01$) corresponding to the "long-before-short" preference. The authors, hence, claim that Persian is another verb-final language that like Japanese displays an effect of relative length corresponding to "long-before-short". Apart from being verb-final, Persian shares all other properties singled out by Yamashita & Chang (2001) in their account of "long-before-short" in Japanese. Persian, like Japanese and contrary to English, displays a fairly free word order and does not require all arguments to be overtly realized. Consequently, following Yamashita & Chang (2001), assuming that longer constituents are lexically richer and hence more salient and conceptually accessible, Faghiri & Samvelian (2014) attribute this ordering preference in Persian to the more important influence of conceptual factors, comparing to form-related ones, in ordering preferences in the preverbal domain in this language.[13]

They note that this preference can be integrated in the continuum established on the basis of the degree of determination of the DO, given that it allows to capture some of the variation observed for the DOs in the middle of the hierarchy. In the case of intermediate DOs, lexical richness contributes to the accessibility of the DO and hence a relatively more salient DO would be located higher in the continuum and therefore is more likely to be separated from the verb, whereas on the two extremities, that is, in the case of marked and bare DOs, the nature of the DO

[13]Note that since Persian is not a head-final language like Japanese, EIC/MiD (Hawkins, 1994, 2004) fails to provide adequate predictions for Persian, as illustrated by Faghiri & Samvelian (2014).

determines its preferred position regardless of relative length.

5 Experimental study

The conclusions of the corpus study conducted by Faghiri & Samvelian (2014) contradicts the broadly accepted view of the relative order between the DO and the IO in Persian, that is, the DOM criterion. According to this view in an unmarked (neutral) word order, *rā*-marked DOs and non-*rā*-marked DOs precede the IO. Lambrecht (1996) highlights that unmarked word orders are not specified for a particular discourse function and can be used in any information structure. Therefore, having more distributional freedom, a neutral word order is the word order that has a greater overall frequency of occurrence. Yet, in the case of indefinite (non-*rā*-marked) DOs, for which the neutral word order is suppose to be IO-DO-V, the inverse order is significantly more frequent in Faghiri & Samvelian's (2014) corpus data.

One could argue that this discrepancy may be of stylistic nature due to the fact that the data is extracted from a journalistic corpus. Therefore, we have run a web-based questionnaire to study the ordering preference of indefinite non-*rā*-marked DOs in a controlled experiment. We designed a sentence completion experiment in order to obtain the preference of speakers for alternative word orders. In this experiment, besides the choice of the order between the two arguments of the verb, the task also required to make a choice between two given possibilities - formally identical but lexically different - for the theme argument. The idea was to bring the attention of the participants to the meaning of the sentence rather than to its form to avoid strategic responses. As mentioned previously, this experiment only included indefinite DOs. The relative length and givenness of the two objects were manipulated following a 2x2 design.

Givenness (or newness) in discourse, that is, the information status, is one highly discussed factor in constituent ordering preferences (e.g. Gundel, 1988; Arnold et al., 2000; Bresnan et al., 2007). In the corpus study of Faghiri & Samvelian (2014) the data was not annotated for the information status and thus the effect of the information structure could not be tested properly.[14] Indefinite DOs are by definition discourse new. As we have mentioned in section 2.3, a discourse given NP in the DO position in Persian is always *rā*-marked. Therefore, it is not possible to manipulate the givenness of an indefinite DO (on a dual discourse-given vs. discourse-new basis). Thus, in this experiment, we manipulated the relative givenness by manipulating the information status of the IO, with two conditions : discourse-new vs. discourse given. In the discourse-given condition, the IO is mentioned in the preamble, whereas in the discourse-new condition there is no

[14]Indeed, DO types are defined on the basis of the degree of determination of the NP and reflect the information status of the DO, more precisely, its referential givenness (see Gundel et al., 1993), to some extent. Yet, the corpus study did not allow to investigate the proper effect of the information structure, that is, independent of grammatical roles.

mention of the IO previously.

With regards to the relative length between the DO and the IO two conditions were defined : DO $>$ IO and DO $<$ IO. We manipulated the length by attaching a modifier ranging from 5 to 10 syllables, with an average of 7 syllables. In the case of the DO, we added adjectives, and in the case of the IO, we added a relative clause. See table 5.1.3 an example of a target item used in the experiment.

5.1 Method

5.1.1 Material

We constructed twenty experimental items in four conditions each. Every item contained a preamble and a target sentence constructed with a ditransitive verb following a $DO_{[-animate]}$ - $IO_{[+human]}$ pattern. We used seven semantically different ditransitive verbs (*baxšidan* 'to donate', *dādan* 'to give', *ferestādan* 'to send', *foruxtan* 'to sell', *gereftan* 'to take', *xaridan* 'to buy', and *xorāndan* 'to feed') implying a variety of prepositions : *be* 'to', *az* 'from', and *barāye* 'for'.[15] For each target sentence two formally identical versions, that is, having the same length and construction but different lexically, of the theme argument were prepared. The experimental items were combined with thirty fillers. Four lists were created according to a Latin Square design.

5.1.2 Procedure

The questionnaire was conducted via the Internet, on the *Ibex-Farm* platform. The participants were asked to take part in the questionnaire only if they had 15 minutes to spare, without doing anything else in the meantime. They were instructed to read the preamble and the three phrases which followed, and construct a sentence, as natural as possible, and fill in the blanks accordingly using drag-and-drop or copy-paste. The instructions indicated that their reaction time was counted in order to put them under some time pressure. Two training items followed the instructions. Figure 3 provides an example of an item on screen. Note that the three phrases were presented in a randomized order.

5.1.3 Participants

33 native speakers of Persian volunteered to complete the web-based (anonymous) questionnaire. They were reached through social networks.

[15] Note that we were limited in our choice of the verb, given the limited number of simplex verbs in Persian, see section 2.4.

Preamble	**Given IO**	
kuler az kār oftāde bud va moštarihā ebrāze nārezāyati mikardand. ābdārči vaqti e'terāzhā be owj resid... 'The air-conditioner had stopped working and the customers were complaining. When protests reached a peak, the janitor...'		
Preamble	**New IO**	
kuler az kār oftāde bud va havā hamintor garmtar mišod. ābdārči vaqti garmā be owj resid... 'The air-conditioner had stopped working and the weather kept getting warmer. When the heat reached a peak, the janitor...'		
Target sentence : .... *dād.* 'gave'		
	DO < IO	
IO *be moštarihā ke az garmā kalāfe budand* 'to customers frustrated from the heat'	**DO1** *yek livān šarbat* 'a glass of syrup'	**DO2** *yek qāč hendevāne* 'a slice of watermelon'
	DO > IO	
IO *be moštarihā* 'to customers'	**DO1** *yek livān šarbate sekanjebine tagari* 'a glass of icy mint syrup'	**DO2** *yek qāč hendevāneye bedune hasteye xonak* 'a slice of fresh seedless watermelon'

Table 1: Example of an item of the experiment

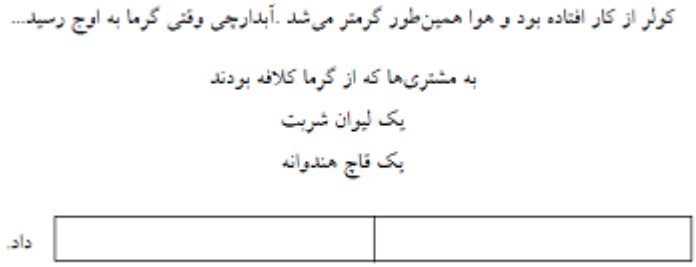

Figure 3: Example of an item on the screen

5.2 Results

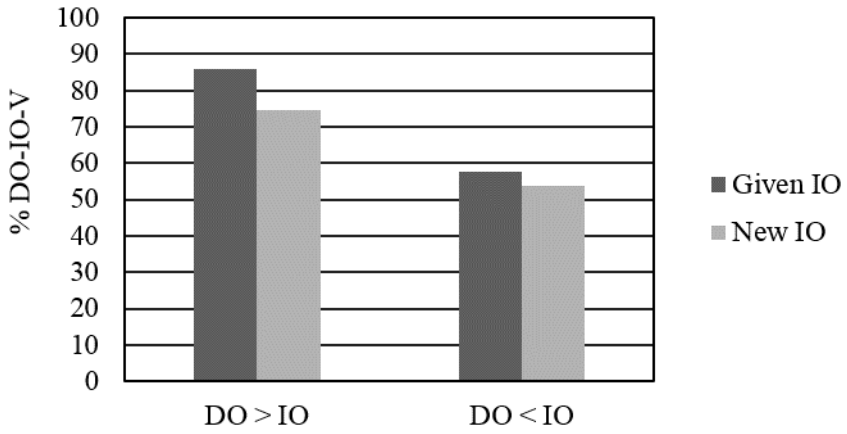

Figure 4: Number of DO-IO-V orders in % by Relative Length and Givenness

The mean proportion of the DO-IO-V order in participants sentences by relative length and givenness is presented in Figure 4. The data confirm an overall preference (68%; χ^2=85.8242, df=1, p<2.2e-16) for the DO-IO-V order. We observe that when the DO is longer that the IO, the preference for the DO-IO-V order is much more stronger than when the DO is smaller than the DO (80.3% vs. 55.7%; χ^2=44.5857, df=1, p=2.435e-11). Surprisingly, when the IO is given, the preference for the DO-IO-V order is stronger than when the IO is new; the difference however is relatively small (71.8% vs. 64.2%; χ^2=4.0127, df=1, p=0.04516). Note that the DO-IO-V order remains above average in all conditions.

To analyses the results statistically, we fitted a mixed-effect logistic model (Agresti, 2007), predicting the relative order between the DO and the IO by the two experimental factors, givenness of the IO and relative length, as fixed effects.

Random effects:				
	Variance	Std. Dev.		
SUBJ (Intercept)	0.7509	0.8666		
ITEM (Intercept)	0.2390	0.4889		
VERB (Intercept)	0.1418	0.3766		
Number of obs: 660, groups: SUBJ, 33; ITEM, 20; VERB, 7				
Fixed effects:				
	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.3909	0.2909	1.344	0.1790
GIVENNESS=IO-NEW	-0.1585	0.24205	-0.655	0.5124
REL-LEN=DO>IO	1.7499	0.2991	5.851	4.9e-09
IO-NEW:DO>IO	-0.7441	0.3916	-1.900	0.0574
Correlation of Fixed Effects:				
	(Intercept)	IO-NEW	DO>IO	
IO-NEW	-0.415			
DO>IO	-0.338	0.402		
IO-NEW:DO>IO	0.256	-0.616	-0.751	
N.B. Success corresponds to ORDER=DO-IO				

Table 2: Results of logistic mixed-effect model

We included subject item and verb as random intercepts,[16] in order to account for inter-subject and inter-item variation in the data, as well as the lexical bias of the verbal lemma. The results of the model are given in Table 5.2.[17]

The analysis shows that relative length has a significant effect ($p < 0.001$) corresponding to the "long-before-short" preference. Givenness of the IO, however, does not turn out to have a significant main effect ($p > 0.5$). There is, nevertheless, a marginal interaction between the two variables ($p < 0.1$). More experiments will be needed in order to pin down this interaction. Note that it may not be surprising that the givenness of the IO does not play a significant role in determining the relative order between the two objects, given the fact that on the hierarchy of the grammatical roles Keenan & Comrie (1977) the IO occupy a lower position than the DO.

[16]The maximal model also included main and interaction slopes for all random effects, but the analysis of variance (ANOVA) showed no significant advantage comparing to the simpler model we present here (χ^2=24.88, df=27, p=0.5812).

[17]These results are fully consistent with (actually nearly identical to) a former experiment run with 60 subjects (no overlaps), in which, due to a script error, each subject had completed a selection of 16 items out of 20.

6 A Unified Account

We propose a unified account of the relative order between the DO and the IO in Persian, on the basis of the conceptual accessibility, grounded in insights provided by Yamashita & Chang (2001) in their account of the "long-before-short" preference in Japanese. Recall that according to incremental models of sentence production, the linear order of constituents is related to their accessibility, in the formal as well as the conceptual arena. In the conceptual arena there is a tendency to place more conceptually accessible constituents, that is, discourse-given, salient, animate, etc., earlier in the sentence (e.g. Bock, 1982; Kempen & Harbusch, 2003)

The empirical evidence presented in this paper shows that the relative order between the DO and the IO in the preverbal domain in Persian depends on two factors. In the first place, the degree of determination of the DO and, in the second place, its length. Indeed, these two independent factors contribute both to the conceptual accessibility of the DO.

1. For an NP in the DO position in Persian, one can safely assume that *rā*-markedness, which corresponds to the highest degree of determination, corresponds also to the highest degree of discourse givenness (and/or prominence). The lowest degree can also be safely assumed to correspond to bare (and bare-modified for that matter) DOs. Indefinite DOs occupy an intermediate position. Hence, the continuum established on the basis of the degree of determination corresponds to a hierarchy of discourse givenness and/or prominence. A factor that contributes to the conceptual accessibility of a constituent.

2. As argued by Yamashita & Chang (2001), longer constituents, containing extra lexical material, are semantically richer, and hence are more salient, that is, more (conceptually) accessible, than shorter ones.

Now let us take a closer look into the ordering preferences for different types of DO with respect to length:

- Marked DOs, uncontroversially, strongly prefer to the DO-IO order, and bare DOs strongly prefer the IO-DO order, regardless of length.

- Indefinite DOs, our study has confirmed, present a moderate preference for the DO-IO order, which increases significantly for longer DOs.

- Bare-modified DOs can be viewed as longer counterparts of bare DOs. They are lexically richer and therefore, even though they display the same degree of discourse givenness as bare DOs, are more salient than the latter and hence conceptually more accessible. The corpus data have showed a rather moderate preference of these DOs for the IO-DO order (comparing to bare DOs), which decreases for longer DOs, see Figure 2 above.

Consequently, to account for these ordering preference as a whole, we suggest to establish a continuum on the basis of the increasing degree of conceptual accessibility - combining discourse givenness/prominence and lexical salience - of the DO, from the strong preference of bare DOs for the IO-DO-V order to the strong preference of *rā*-marked DOs for the DO-IO-V order.

7 Conclusion

In this paper, we have presented experimental data on the relative order between the DO and the IO in Persian to follow up on the corpus study conducted by Faghiri & Samvelian (2014). These findings have a twofold interest, one *vis-à-vis* the existing hypothesis for Persian, and second, with respect to the effect of the relative length cross-linguistically.

1. The position of the DO does not exclusively depend on its markedness. Ordering preferences of verbal complements in ditransitive constructions reflect a continuum on the basis of the degree of determination of the DO rather than a categorical behavior depending on its markedness. This fact contradicts a dual syntactic position hypothesis for the DO, as claimed by some theoretical studies (e.g. Karimi, 2003).

2. Ordering preferences of verbal complements in ditransitive constructions show a significant effect of relative length corresponding to the "long-before-short" principle. Supporting mirror-image preferences in OV and VO languages. Note that, as Faghiri & Samvelian (2014) argued, Persian data is not predicted by the EIC/MiD principle (Hawkins, 1994, 2004).

Reinforcing the hypotheses formulated in Faghiri & Samvelian (2014), we have proposed a unified account of the position of the DO based on its conceptual accessibility. We have combined discourse givenness (and/or prominence), on the one hand, and lexical richness on the other hand, both assumed to contribute to the conceptual accessibility of a constituent. Thus, we have provided an empirically valid account that not only predicts the preferred position of different types of DOs, but also accounts for the variation in the strength of these preferences.

The experimental study we presented in this paper is the first of a series of experiments we are undertaking to study ordering preferences between the DO and the IO in the preverbal domain in Persian. Namely, similar experiments for bare-modified DOs and experiments to test the gradual nature of ordering preferences depending on the DO type, are underway.

One open issue remaining is the role of the subject. As a matter of fact, in the literature that discuss word order variations in the preverbal domain in Persian, the position of the subject is rarely discussed, most probably because the neutral SOV word order is uncontroversial. The relative order of objects, however, is a long lasting debate, mainly because of the DOM in Persian. Nevertheless, it is

crucial to include the subject in studies on ordering preferences in the preverbal domain in Persian, as it is also the case with Yamashita & Chang (2001). More precisely, it is interesting to see to what extent a highly (conceptually) accessible DO, that is, for example, an animate *rā*-marked DO, is likely to win the competition for the initial position of the sentence over the subject. In future research, we are taking on ordering preferences between the subject and the direct object in transitive constructions, as well as between all the three constituents in ditransitive constructions.

References

Agresti, Alan. 2007. *An introduction to categorical data analysis*. Wiley-Interscience.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3). 435–483.

Arnold, Jennifer E., Thomas Wasow, Anthony Losongco & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language* 76(1). 28–55.

Bock, J. Kathryn. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89(1). 1–47.

Bock, J. Kathryn & Willem Levelt. 1994. Language production: Grammatical encoding. In Morton A. Gernsbacher (ed.), *Handbook of psycholinguistics*, 945–984. New York: Academic Press.

Bossong, Georg. 1985. *Empirische Universalienforschung: differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen : Gunter Narr Verlag.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Boume Gerlof, Irene Kraemer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94.

Chang, Franklin. 2009. Learning to order words: A connectionist model of heavy {NP} shift and accessibility effects in japanese and english. *Journal of Memory and Language* 61(3). 374 – 397.

Choi, Hye-Won. 2007. Length and order: A corpus study of Korean dative-accusative construction. *Discourse and Cognition* 14(3). 207–227.

Corbett, Greville G. 2000. *Number*. Cambridge University Press.

Dabir-Moghaddam, Mohammad. 1992. On the (in) dependence of syntax and pragmatics: Evidence from the postposition *-ra* in Persian. In *Cooperating with written texts: The pragmatics and comprehension of written texts*, 549–573. Mouton de Gruyter.

Faghiri, Pegah. 2010. *La morphologie du pluriel nominal du persan d'après la théorie whole word morphology*. Université de Montréal MA thesis.

Faghiri, Pegah & Pollet Samvelian. 2014. Constituent ordering in Persian and the weight factor. In Christopher Pinon (ed.), *Empirical issues in syntax and semantics 10 (EISS10)*, In press.

Ganjavi, Shadi. 2007. *Direct objects in persian*. University of Southern California.

Garrett, Merrill F. 1980. Levels of processing in sentence production. *Language production* 1. 177–220.

Ghomeshi, Jila. 1997. Topics in persian vps. *Lingua* 102(2—3). 133–167.

Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia linguistica* 57(2). 47–74.

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

Gundel, Jeanette K. 1988. Universals of topic-comment structure. In M. Hammond, E.A. Moravcsik & J.R. Wirth (eds.), *Studies in syntactic typology*, 209–239. John Benjamins.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

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

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford University Press.

Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Blackwell Publishing.

Keenan, Edward L & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic inquiry* 8(1). 63–99.

Kempen, Gerard & Karin Harbusch. 2003. Word order scrambling as a consequence of incremental sentence production. In Holden Härtl & Heike Tappe (eds.), *Mediating between concepts and grammar*, 141–64. Mouton De Gruyter.

Lambrecht, Knud. 1996. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.

Lazard, Gilbert. 1982. Le morphème *rā* en persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris* 77(1). 177–208.

Lazard, Gilbert, Yann Richard, Rokhsareh Hechmati & Pollet Samvelian. 2006. *Grammaire du persan contemporain*. Institut français de recherche en Iran.

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.

Roberts, John R, Behrooz Barjasteh Delforooz & Carina Jahani. 2009. *A study of persian discourse structure*. Acta Universitatis Upsaliensis.

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

Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian ezafe. *Journal of Linguistics* 43(3). 605–645.

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

Samvelian, Pollet & Pegah Faghiri. 2013. Introducing PersPred, a syntactic and semantic database for Persian complex predicates. In *Proceedings of the workshop on multiword expressions*, 11–20. Atlanta, Georgia, USA.

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

Stallings, Lynne M., Padraig G. O'seaghdha & Maryellen C. MacDonald. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journtal of Memory and Language* 39(3). 392 – 417.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9(01). 81–105.

Wasow, Thomas. 2002. *Postverbal behavior*. CSLI.

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