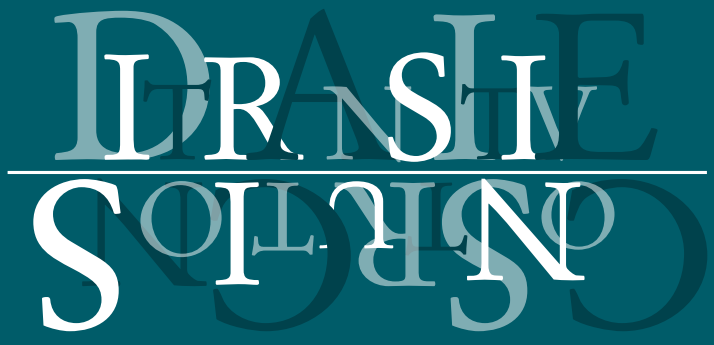
# Is there a canonical order in Persian ditransitive constructions? Corpus based and experimental studies

Pegah Faghiri, Pollet Samvelian, Barbara Hemforth

# Ditransitive Constructions in a Cross-Linguistic Perspective

Edited by Agnes Korn and Andrej Malchukov



Reichert

# Ditransitive Constructions in a Cross-Linguistic Perspective

Edited by
Agnes Korn and Andrej Malchukov

Reichert Verlag Wiesbaden 2018

Printed with the financial support of the
Department of Linguistics and Philology of Uppsala University.

**Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Gedruckt auf säurefreiem Papier
(alterungsbeständig – pH7, neutral)


www.reichert-verlag.de
ISBN: 978-3-95490-373-3

# Table of Contents

Introduction and Acknowledgements ..... 7

Typological remarks on "internal" beneficiaries
and the benefactive-possessive convergence
ANDREJ MALCHUKOV ..... 13

Double accusative constructions and ditransitives in Ancient Greek
SILVIA LURAGHI & CHIARA ZANCHI ..... 25

Ditransitive constructions in Tocharian B
PAUL WIDMER ..... 49

Ditransitive constructions in Latin: Competition of paradigmatic dimensions
ELENA ZHELTOVA ..... 63

Overt and covert ditransitivity in Romance:
From double object constructions to case alternations
ANNA PINEDA ..... 77

Constructing ditransitivity in literary Kurmanji: Some fuzzy issues
ANNETTE HERKENRATH ..... 95

Ditransitive constructions in Tāti
MOHAMMAD RASEKH-MAHAND & RAHELEH IZADIFAR ..... 119

Ditransitivity constructions in three Caspian dialects
GUITI SHOKRI ..... 135

Ditransitive constructions in three Balochi dialects from a typological perspective
MARYAM NOURZAEI & THOMAS JÜGEL ..... 147

Is there a canonical order in Persian ditransitive constructions?
Corpus-based and experimental studies
PEGAH FAGHIRI, POLLET SAMVELIAN & BARBARA HEMFORTH ..... 165

Ditransitive constructions in Mansi dialects
BERNADETT BÍRÓ & KATALIN SIPŐCZ ..... 187

Ditransitive constructions in selected Niger-Congo languages in a typological perspective
ALEXANDER ZHELTOV ..... 207

**Appendix**
List of contributors ..... 227

Maps:

Location of the Mazandarani varieties studied ..... 136
Balochi dialects ..... 148
Distribution of Mansi dialects ..... 189

# Is there a canonical order in Persian ditransitive constructions? Corpus-based and experimental studies

Pegah Faghiri, Pollet Samvelian & Barbara Hemforth

## Abstract

The canonical word order in ditransitive sentences constitutes an essential argument in the development of the prevailing existing analyses of the VP in Persian, an SOV language with flexible word order and Differential Object Marking. Adopting a quantitative approach to word order variations, FAGHIRI & SAMVELIAN (2014) and FAGHIRI et al. (2014) have respectively provided corpus and experimental data that undermine the broadly admitted assumption of the canonical order between the two objects in ditransitive sentences. In this paper, we provide complementary corpus and experimental data on this issue; in particular, focusing on the extent of word order variations in ditransitive sentences, we report comparable overall rates on word order variations in transitive sentences, as a benchmark. Our findings highlight that the relative order between the two objects displays much more variation than generally assumed, in such manner that positing a canonical order, similar to SOV, is misleading for ditransitive sentences. We argue instead that the relative order between direct and indirect objects is a matter of soft constraints and results from the interaction of a set of functional factors, in conformity with the general "salient-first" preference.

## 1. Introduction

The goal of this paper is to discuss the relevance of positing a canonical order for ditransitive sentences in Persian. The canonical word order in ditransitive sentences is much less discussed in the general literature on canonical or basic word order than the linear order between basic constituents, that is, in transitive sentences, subject, object and verb. However, the "canonical" relative order between the direct object (DO) and the indirect object (IO) has received a great deal of attention in studies on Persian syntax. This is principally due to Persian's Differential Object Marking (DOM), cf. Section 1.1. In fact, the syntactic position of marked vs. unmarked DOs is at the heart of the prevailing analyses of the Persian VP, mainly developed in the generative framework. These analyses build on a broadly admitted assumption on the canonical or, as formulated in more recent studies, the neutral (unmarked) relative order between direct and indirect objects in ditransitive sentences, hereafter the DOM criterion, cf. Section 1.2.[1]

[1] We would like to thank the audience of the Workshop on Ditransitive Constructions in a Cross-linguistic Perspective, adjacent to the SWL 6 Conference at the University of Pavia (Italy) and the editors of this volume, Agnes Korn and Andrej Malchukov. We would also like to thank two anonymous reviewers for their helpful comments on the first version of this paper.
This work is partly 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).

According to this assumption, in a canonical ditransitive sentence, marked DOs precede, while unmarked DOs follow the IO. Recent empirical studies, including corpus-based (FAGHIRI & SAMVELIAN 2014) and experimental studies (FAGHIRI et al. 2014), have undermined this hypothesis by clearly showing that it provides wrong predictions for a (non-negligible) part of unmarked DOs. To this end, rather than the usual marked vs. unmarked classification, the authors have defined a more fine-grained classification of Persian DOs, based on the degree of determination of the DO, briefly presented in Section 1.1.[2]

Different definitions are available in the literature for canonical word order (for a review, cf. e.g. BRODY, 1984; SIEWIERSKA, 1988; DRYER, 2007). One of the most agreed upon definitions stipulates that in a given language, the basic word order is the order in which constituents appear in the least pragmatically and stylistically marked or neutral declarative sentences. This explains why many authors employ canonical word order and neutral/unmarked word order as synonyms. Pragmatic (un)markedness is, however, not a straightforward notion and identifying the neutral word order among semantically, or truth-conditionally, equivalent word orders is not (always) self-evident. Frequency is one single criterion upon which most authors rely to identify the dominant word order (if any) among competing orders. In fact, for most authors, frequency coincides with, or is a byproduct of, pragmatic neutrality. In other words, the least (pragmatically) marked word order will also be the most frequent one, by virtue of having more distributional freedom (cf. e.g. LAMBRECHT, 1996). It is important to note that the use of frequency as a diagnostic for canonical word order is not uncontroversial. The two main problems pointed out in the literature (cf. e.g. SIEWIERSKA 1988; cf. DRYER 1995 for a discussion) are the following:
1) in many languages sentences with full NPs are infrequent so that ordering preferences cannot easily be obtained due to sparse data problems; and 2) word order frequencies may vary in a same language from one type of corpus to another. Thus, the way frequencies were obtained and compared in many earlier studies suffer methodological shortcomings. However, these (justified) problems are solvable if the methodological requirements of quantitative analysis are met. Frequencies of competing orders need to be compared in the most controlled and unbiased manner possible, meaning controlling for confounding factors as much as possible. Indeed, the quantitative approach, provided methodological standards are respected, remains the most objective and reliable way to identify the canonical word order, among competing word orders.

Importantly, we do not claim that a quantitative analysis needs to be pursued for each and every linguistic hypothesis. Naturally, in some cases intuitions are so strong and uncontroversial that they can be regarded as robust and highly reliable by themselves. Few researchers would undertake a quantitative study to verify if the canonical word order in English is effectively SVO. The same is true for the SOV in Persian, which without any doubts is the canonical word order in this language.[3] Nevertheless, in some other cases,

[2] For a more detailed presentation, see SAMVELIAN & FAGHIRI (2014: 216-219).

[3] It is however important to bear in mind the crucial difference between these two languages. English is a fixed word order language in which grammatical relations depend on the word order. Persian, on the contrary, is a (relatively) free word order language in which the six possible basic word orders are possible and (basic) grammatical relations do not depend on the word order.

variations are more abundant and the difference between available alternatives more subtle, hence intuition is much less reliable. This is, for example, the case for the well-known dative alternation in English, for which only quantitative studies have succeeded in providing a reasonably precise picture of factors determining ordering preferences (e.g. COLLINS 1995; WASOW 2002; BRESNAN et al. 2007). In particular, these studies have clearly highlighted the effect of functional factors, such as the relative length, definiteness, discourse givenness and animacy.

The difference between these two situations is reminiscent of "hard" vs. "soft" constraints, in the sense of SORACE & KELLER (2005). For instance, while the SVO order in English results from a "hard" or grammatical constraint on the phrase structure of this language, the choice between dative or double object constructions results from the interaction of a set of "soft" constraints or preferences. The relative order between the DO and IO in Persian seems to be of a similar nature. Indeed, the above-mentioned quantitative studies report more word order variations than generally (and theoretically) assumed and furthermore show that relative length (in number of words) has a significant effect on the linear order between the two objects.

In the present study, we focus on the extent of word order variations in ditransitive sentences, in order to determine whether it is relevant to posit a canonical schema, or a "hard" constraint. To this end, as a benchmark, we compare the distribution of relative order between the DO and the IO with the distribution of word order in transitive sentences, in corpus as well as experimental data. Building on these data, we argue that whereas SOV clearly emerges as the canonical order of basic constituents in this language, there is no empirical ground to identify any order as the dominant word order in ditransitive sentences. In other words, while the SOV order operates, to some extent (see footnote 3) as a "hard" constraint on the relative order between the subject and the object in an unmarked sentence, the relative order between the DO and the IO results from the interaction of a set of "soft" constraints.

This paper is organized as follows. In the reminder of this section, we discuss different realizations of the DO in Persian (Section 1.1.) and present the DOM criterion (Section 1.2.). Section 2. provides the corpus analyses, and Section 3. the experimental data. In Section 4., we propose an account of ordering preferences between the DO and the IO in terms of interacting functional universals. Finally, Section 5. concludes the paper.

### 1.1. DOM and Different Realizations of the DO in Persian

DOM,[4] realized by the enclitic *=rā*, is a well-known feature of Persian, yet object of ongoing controversial debate. Indeed, no uncontroversial straightforward account of the latter is available in the literature (see SAMVELIAN 2018 for a review). Traditionally, *rā*-markedness is related to definiteness. Indeed, a definite DO is always marked by the enclitic *=rā,* ex. (1a).

[4] Coined by BOSSONG (1985), DOM denotes the property of some languages with overt case-marking of some but not all direct objects depending on semantic and pragmatic features, see also AISSEN (2003).

In (formal) Persian there is no overt marker for definiteness. Definite NPs can be formed either by different definite determiners, like demonstratives, ex. *in medād* 'this pencil', or by no overt determination, *medād* 'the pencil'. In the DO position, an NP lacking *=rā*, and carrying no determination or quantifications, like *medād* or *medād=e qermez* in (1b), will necessarily receive a "bare noun" reading, that is, a nonspecific (existential or a kind-level/generic) reading. Bare nouns are underspecified for number in Persian and can consequently yield a mass reading (even if countable).

(1a) *Mahsā* *(in) medād=rā* *xarid*
Mahsa (this) pencil=DOM bought
'Mahsa bought (this/) the pencil.'

(1b) *Mahsā* *medād(=e qermez)* *xarid*
Mahsa pencil(=EZ[5] red) bought
'Mahsa bought (red) pencils/a (red) pencil.'

Meanwhile, DOM is far more complex and cannot be captured by a binary ±definite feature, namely because *=rā* can co-occur with the indefinite determination, as in (2). Contrary to definitess, indefiniteness is overtly marked in Persian, by the enclitic *=i*, ex. *medād=i*, the cardinal *yek,* ex. *yek medād*, or the combination of both, ex. *yek medād=i* 'a pencil'. Naturally, indefinite NPs are also formed by different indefinite quantifiers (the noun remains in the singular form even when it denotes more than one entity), *čand medād* 'a few pencils'. Indefinite DOs are always specified for number and have an existential reading. An indefinite DO carrying *=rā* will receive a specific reading, as in (2b). This fact has led some scholars to analyze *=rā* as triggered by specificity instead of definiteness (e.g. BROWNE 1970; BROWNING & KARIMI 1994; KARIMI 2003). Some others treat *=rā* as a marker of topicality (e.g. DABIR-MOGHADDAM 1992), since *rā*-marked DOs can have a generic, that is, a nonspecific, reading.

(2a) *Mahsā* *(yek)* *medād=i* *xarid*
Mahsa (a) pencil=INDF bought
'Mahsa bought a pencil.'

(2b) *Mahsā* *(yek)* *medād=i=rā* *xarid*
Mahsa (a) pencil=INDF=DOM bought
'Mahsa bought a (certain) pencil.'

Building on larger sets of data, several scholars point out that no binary feature is capable of providing a clear-cut account of DOM in Persian (cf. e.g. LAZARD 1982; GHOMESHI 1997). They argue that the latter can only be accounted for on the basis of a set of properties, such as definiteness, animacy, individuation, totality, topic-hood, that contribute to the high transitivity in the sense of HOPPER & THOMPSON (1980).

What we retain from these observations and discussions is that in a given context the more (discourse) salient a DO the more it is likely to be marked with *=rā*. Hence, among different possible realizations of the DO that we consider in this study, that is, *rā*-marked,

[5] The enclitic *=(y)e*, the *Ezafe*, links the head noun to its modifiers and to the possessor NP (cf. SAMVELIAN 2007).

ex. (3a), indefinite (non-*rā*-marked), ex. (3b) and bare nouns – separating single-word bare nouns, ex. (3d), from bare nouns enriched with (restrictive) adjuncts, ex. (3c), *rā*-marked DOs are the most discourse accessible. Furthermore, we assume that bare DOs appear at the lowest position on this continuum, that is, they are the least salient discourse accessible type of DO in Persian. Indefinite (non-*rā*-marked) DOs appear in between.

(3) a. *rā*-marked DOs: *(in) medād=rā, (yek) medād(=i)=rā*, etc.
b. Indefinite (non-*rā*-marked) DOs: *yek medād(=i), čand medād*, etc.
c. Bare-modified DOs: *medād=e qermez*, etc.
d. Bare DOs: *medād*

1.2. The DOM criterion

It is generally assumed that the relative order between the direct object (DO) and the indirect object (IO) in Persian depends on DOM (e.g. KARIMI 1990; MAHOOTIAN 1997; ROBERTS et al. 2009). Following FAGHIRI & SAMVELIAN (2014: 221), we refer to this generalization as the DOM criterion, illustrated by the schema in (4). According to this hypothesis, non-*rā*-marked DOs follow the IO, ex. (5a), while *rā*-marked DOs precede it, ex. (5b).

(4) The DOM Criterion:
a. (S) IO DO V
b. (S) DO=*rā* IO V

It should be noted that the generalization in (4) is not always formulated in terms of *rā*-markedness. Authors usually state this assumption in terms of properties they assume to trigger DOM, such as definiteness, specificity or topicality. We prefer to rely only on the formal criterion of *rā*-markedness and avoid reference to any semantic and/or pragmatic features, which moreover remain controversial and difficult to pin down.

(5a) *Ali be Minu āb dād*
Ali to Mina water gave
'Ali gave water to Minu.' (MAHOOTIAN 1997: 6)

(5b) *Ali āb=rā be Minu dād*
Ali water=DOM to Minu gave
'Ali gave the water to Minu.' (MAHOOTIAN 1997: 7)

This polarized view of the neutral word order in ditransitive constructions has led many scholars, mainly in the generative framework, to suggest a hierarchical view of the Persian VP (KARIMI 1990, 2003, 2005; BROWNING & KARIMI 1994; GHOMESHI 1997; GANJAVI 2007). In this view, hereafter the "Two Object Position Hypothesis" (TOPH), (KARIMI 2003: 105), *rā*-marked DOs occupy a higher structural position than their non-*rā*-marked counterparts, as illustrated by schema (6). This view of the phrase structure of Persian's VP is the dominant view in the literature (see SAMVELIAN, 2001, BONAMI & SAMVELIAN 2015, and FAGHIRI & SAMVELIAN 2016, for an alternative view).

(6) a. $[_{VP}\ DP_{[+r\bar{a}]}\ [_{V'}\ PP\ V]]$
b. $[_{VP}\ [_{V'}\ PP\ [_{V'}\ DP_{[-r\bar{a}]}\ V]]]$ (adopted from KARIMI 2003:105)

This structure obviously provides a straightforward account for the canonical word order presented above. Furthermore, authors claim that it provides a syntactic explanation for the tight semantic bond that exists between the verb and its non-*rā*-marked DOs, contrary to its *rā*-marked DOs (cf. e.g. KARIMI 2003). This bond is argued to be reminiscent of the relationship between a bare noun, that is a non-*rā*-marked DO *par excellence*, and a verb in a noun-verb complex predicate (CPr).[6]

## 2. Corpus Data

FAGHIRI & SAMVELIAN (2014) present the first corpus-based study on the relative order between the DO and the IO in Persian in order to evaluate the DOM criterion (see Section 1.2.). To this end, the authors study the distribution of word order in an unbiased sample of sentences with ditransitive (verb final) patterns extracted from the Bijankhan corpus, an open access corpus[7] of 2.6 million tokens manually tagged for POS, extracted from the Iranian *Hamshahri* daily.

As a benchmark enabling us to assess the extent of word order variations in ditransitive constructions observed in their data, we study the relative order between the subject and the direct object in the same corpus.[8] In other words, our goal is to find out whether the variation reported by FAGHIRI & SAMVELIAN (2014) for ditransitive constructions aligns with the variation we would observe for transitive constructions. Before presenting our data on the latter, let us first present a brief summary of FAGHIRI & SAMVELIAN (2014)'s study.

### 2.1. Word Order Variation in Ditransitive Sentences

FAGHIRI & SAMVELIAN (2014) constituted a dataset containing 905 sentences with ditransitive verb final patterns, that is, NP-PP-V or PP-NP-V, which they manually filtered out of an unbiased sample extracted from the Bijankhan corpus (for details, see FAGHIRI & SAMVELIAN 2014: 222-224). Their dataset contains 1) 424 *rā*-marked, 2) 144 indefinite, 3) 66 bare-modified and 4) 271 bare DOs.

*Rā*-marked DOs, as predicted by the DOM criterion, display a very strong preference (95%) for the position separated from the verb. However, their unmarked counterparts display more variation. Bare DOs have a strong preference (84.2%) for verb adjacency but bare-modified DOs have a more moderate preference (66.7%) for this position. Indefinite DOs, contrary to the DOM criterion, have a clear preference (77%) for the non verb-adjacent position.

---

[6] Due to its limited number of simplex verbs, Persian is characterized by the significant presence of complex predicates (CPr), that is, the combination of a non-verbal element (mostly a bare noun) with a verb that behaves as one semantic unit, ex. *harf zadan* 'to talk (Lit. speech (to) hit)'. For a detailed presentation of CPrs in Persian as well as an alternative view of the latter, see e.g. SAMVELIAN (2012), SAMVELIAN & FAGHIRI (2014).

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

[8] To be exact, these studies are both carried out on a version of the Bijankhan corpus in which verbs are lemmatized (see FAGHIRI 2016 for details). The initial version of this corpus does not contain syntactic annotation, nor is it lemmatized or delimited for sentences. Finite verbs are used as cue for the automatic extraction of potential sentences, skimmed off manually to obtain the datasets from which the frequencies are calculated.

In line with studies on word order variation across languages, FAGHIRI & SAMVELIAN (2014) also focus on the effct of the grammatical weight on ordering preferences, in terms of differences in relative length between constituents counted in number of words. The authors fit a mixed-effect regression model to a subset of their dataset, excluding *rā*-marked and bare DOs, with DO type and relative length as fixed effects and verbal lemma as random effect. Their results confirm the DO-IO-V order for indefinite DOs and the inverse for bare-modified ones (positive coefficient vs. negative coefficient, p-values < 0.001) and show a significant effect (p-value < 0.001) of the relative length corresponding to the "long-before-short" preference (for details, see FAGHIRI & SAMVELIAN 2014: 224).

## 2.2. Word order Variations in Transitive Sentences

To study the distribution of word order in transitive constructions, we elaborated a first dataset upon a sample of 1500 occurrences of the verb *didan* 'to see' – as a typical transitive verb – randomly extracted out of the Bijankhan corpus. We identified the relevant occurrences manually according to the following criteria: We chose all propositions in which the three basic constituents, subject, DO and verb, are realized. We excluded cases in which the DO is a clausal complement and accordingly appears in the postverbal domain. We also excluded highly stylistically marked occurrences such as poetry. We then annotated the remaining 245 occurrences for word order and the DO type. In line with FAGHIRI & SAMVELIAN (2014), we did not exclude occurrences in which the DO and the verb form a complex predicate from our selection.

Table 1: Word order distribution in a random sample of the verb *didan* 'to see' from the Bijankhan Corpus

| DO type | *rā*-marked | Indefinite | Bare-modified | Bare | Total |
|---|---|---|---|---|---|
| SOV | 142 (94%) | 40 (100%) | 8 (100%) | 46 (100%) | 236 (96.4%) |
| OSV | 9 | 0 | 0 | 0 | 9 |
| Total | 151 | 40 | 8 | 46 | 245 |

The distribution of word order in the data, given in Table 1, reveals a very important rate of the SOV order, 96.4%. Interestingly, in this data, only *rā*-marked DOs appear in the OSV order. We have repeated this operation for the verbs *neveštan* 'to write' and *koštan* 'to kill' and observed similar word order distributions (92% and 100% respectively), and found no occurrences of the OSV order with a non *rā*-marked DO.

Upon this observation, we decided to observe word order variations for a sample of *rā*-marked DOs. To this end, we randomly extracted 900 occurrences of *=rā* from the corpus. We then manually identified relevant occurrences for our study, according to the same criteria mentioned above. Our dataset contains 190 occurrences of transitive sentences, with either simplex or complex verbs, of which 177 (93.2%) are in SOV order, 12 (6.3%) in the OSV order and only 1 (0.05%) in the SVO order.

Note that this distribution of word order contrasts sharply with the data on the relative order between the DO and the IO. In order to highlight this divergence, Table 2 provides the

distribution of word order in a sample of occurrences of the verb *dādan* 'to give' to compare with that of the verb *didan* provided above.

Table 2: Word order distribution in a random sample of the verb *dādan* 'to give' from the Bijankhan Corpus

| DO type | *Rā*-marked | Indefinite | Bare-modified | Bare | Total |
|---|---|---|---|---|---|
| DO-IO-V | 62 (86.1%) | 46 (79.3%) | 8 | 11 | 127 (65.1%) |
| IO-DO-V | 11 | 12 | 24 (75%) | 22 (66.7%) | 68 |
| Total | 72 | 58 | 32 | 33 | 195 |

2.3. Conclusions

The distribution of word order in these data clearly allows us to identify the SOV order as the canonical word order in transitive sentences in Persian, which is almost systematically preferred over other competing word orders. On the contrary, the picture is much less clear-cut for the distribution of competing word orders for the relative order between the DO and the IO (in the same corpus). Only the overall rate observed for *rā*-marked DOs, 95% for the DO-IO-V order, reaches the same level as the rate for the SOV order. Non-*rā*-marked DOs, with only 63% for IO-DO-V order, display indeed a great amount of variation. On the other hand, if we oppose bare single-word DOs to all other DO types, the rates amount respectively to 84% and 85% for IO-DO-V and DO-IO. Bareness by it-self, including bare and bare-modified DOs, gains a less important rate 80% (for the IO-DO-V order) and inversely a rather higher rate for non-bare DOs, 90% (for the DO-IO-V order).

## 3. Experiments

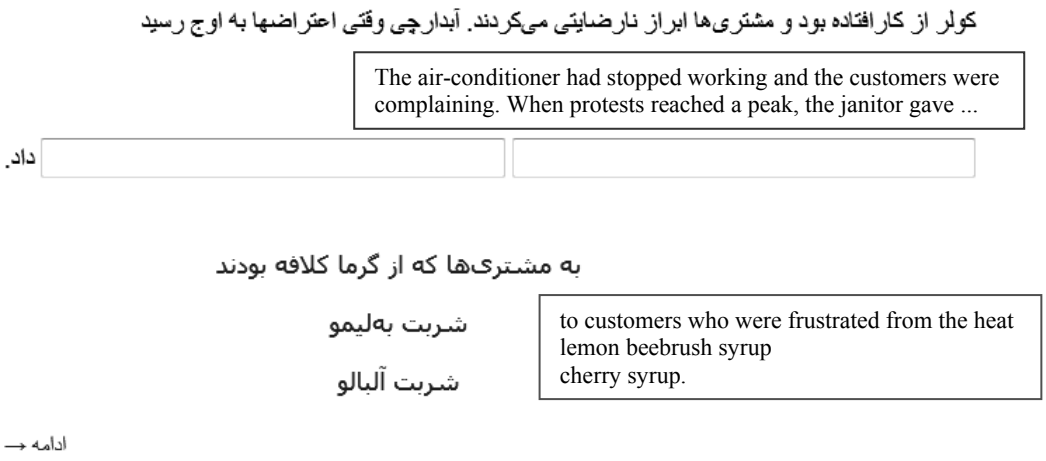
In this section, we present a series of sentence completion experiments carried out in order to get a more precise picture following the conclusions of the corpus analyses. It could be argued that word order distributions reported in these studies may be of stylistic nature given that the data is extracted from a journalistic corpus. The experimental method allows us to collect preferences from native speakers in an indirect and reasonably unbiased and objective manner, contrary to earlier studies in which grammatically judgments are elicited informally from a small number of speakers and with no effort to ensure their objectivity.[9] In these experimental studies, word order preferences of native speakers of Persian are elicited via web-based controlled questionnaires. Sentences, or technically speaking, experimental items, are prepared following Latin Square designs and presented to each participant in a randomized order separated by two distractor items. Participants are recruited through social networks, on a voluntary basis.

In each experiment, participants are asked to complete a sentence, which means, to fill in the blanks, with provided options. Each experimental item (see Figure 3) contains a sentence in which a number of constituents are missing, represented by a context-setting

[9] Previous studies rely on acceptability judgments elicited with almost no methodological concerns.

beginning and the corresponding number of blank boxes, for instance, two blank boxes for the DO and the IO. To complete the sentence, that is, to fill in each blank, participants are provided with a list of possible constituents (presented simultaneously on the screen in a counterbalanced order). The relative order between these constituents (in the final sentence) is left to the participants and constitutes the response or the dependent variable. The list of options always contains one element more than the number of blanks in order to prevent participants from guessing the purpose of the experiment and to push them to concentrate on the content of each sentence and to produce reasonably natural sentences.

Figure 3: Example of an item from Experiment 1




Since corpus data contradict the predictions of the DOM criterion for indefinite (unmarked) DOs, FAGHIRI et al. (2014) conducted a first experimental study to examine the behavior of the latter. Furthermore, they investigated the effect of relative length, as well as givenness on linear order. Their findings converged with the results of the corpus study in: 1) refuting the DOM criterion, and 2) establishing the "long-before-short" preference for Persian. Section 3.1. provides a brief summary of this experiment.

To achieve a well-grounded and comprehensive account of the canonical word order in ditransitive sentences, we have carried out four more experiments. Three experiments are conducted to examine ordering behaviors of all DO types in order to have comparable data for the full pattern: Experiment 1 is the exact replication of FAGHIRI et al. (2014)'s experiment for bare-modified DOs. Experiment 2 studies the behavior of indefinite (unmarked) DOs in comparison to *rā*-marked DOs and Experiment 3 the behavior of bare-modified DOs in comparison to bare single-word ones. Experiment 4 focuses on the relative order in transitive sentences.

It is important to note that the three last experiments are conducted simultaneously via a unique questionnaire.[10] Consequently, we dispose of within subject data allowing us to compare, on one hand, the relative order between the DO and the IO for all DO types, and

on the other hand, the word order distribution in ditransitive vs. transitive sentences. Section 3.6. provides a discussion of the results of these experiments.

### 3.1. Summary of the results of FAGHIRI et al. (2014)'s experiment

In this experiment, FAGHIRI and colleagues focused on indefinite (non-*rā*-marked) DOs. They defined four experimental conditions to manipulate 1) the relative length with DO longer than IO (DO > IO) and IO longer than DO (IO > DO) and 2) the givenness of the IO, with IO given and IO new conditions. All IOs are construed as human beneficiary arguments and DOs as inanimate (concrete) themes, as in (7).[11]

(7)

| | | | | | | |
|---|---|---|---|---|---|---|
| *yek* | *livān* | *šarbat(=e* | *sekanjebin=e* | *tagari)* | *be* | *moštari-hā* |
| a | glass | syrup=EZ | mint=EZ | icy | to | client-PL |
| *(ke* | *az* | *garmā* | *kalāfe* | *bud-and)* | *dād* | |
| that | from | heat | impatient | were-3PL | gave | |

'(S)he gave a glass of (icy mint) syrup to the clients (who were frustrated from the heat).'

Their results show an overall rate of 68% for the DO-IO order, which is consistent with the results of FAGHIRI & SAMVELIAN (2014)'s corpus study, confirming the preference of indefinite DOs to be separated from the verb, contrary to the DOM criterion. Furthermore, they find a significant effect ($p < 0.001$) of relative length, corresponding to a "long-before-short" preference. The rate of the DO-IO order varies in their data from 80.3% for the DO > IO condition to 55.7% for the IO > DO condition (for more details, see FAGHIRI et al. 2014: 231-232).[12]

### 3.2. Experiment 1: Replication of FAGHIRI et al. (2014)'s experiment for bare-modified DOs

#### 3.2.1. Method and material

This experiment is the exact replication of FAGHIRI et al. (2014)'s experiment for bare-modified DOs (instead of indefinite (non-*rā*-marked) DOs), with the only exception that for relative length, we only manipulated the length of the IO and kept the length of the DO constant, yielding IO>DO and IO=DO conditions, as in (8).[13] We used the same material that were used in FAGHIRI et al. (2014)'s experiment and applied only some small modifications to adjust the target items. Twenty-six native speakers of Persian, recruited via social networks, volunteered to take part in this experiment.

(8)

| | | | |
|---|---|---|---|
| *šarbat=e* | *ālbālu* | *be* | *moštari-hā* |
| syrup=EZ | cherry | to | client-PL |

[10] Items of one experiment are used as fillers for the others. 10 extra fillers are also added to the final list in order to make sure that there are always at least two fillers between items of a same experiment.

[11] For a full example of an item in its four conditions, see FAGHIRI et al. (2014: 230).

[12] With respect to the givenness of the IO, surprisingly the rate of DO-IO order is slightly bigger for given IOs than for new IOs (71.8% vs. 64.2%). For a discussion see FAGHIRI (2016: 177-178).

[13] Indeed, it was very challenging to find bare nouns with longer modifiers that would not compromise the naturalness of items.

*(ke* *az* *garmā* *kalāfe* *bud-and)* *dād*
that from heath impatient were-3PL gave
'(S)he gave some cherry syrup to the clients (who were frustrated from the heat).'

3.2.2. Results

Table 4 presents the raw data,[14] Figure 5 summarizes the distribution of the word order with respect to our two experimental factors, and Table 6 presents the results of our mixed-effect logistic regression model including participants and items as random intercepts (in order to account for inter-subject and inter-item variation in the data).[15] The results show an overall rate of 90% for the IO-DO order in conformity with the canonical word order expected for bare DOs. Similar to FAGHIRI et al. (2014)'s result, we find a significant effect of relative length. The rate of IO-DO order varies from 85.5% for the DO=IO condition to 94.2% for the IO > DO condition. Givenness does however not show a main effect nor interact with relative length.[16]

Table 4: Raw data and percentages for Experiment 1

| Conds. | DO-IO | IO-DO | % of IO-DO |
|---|---|---|---|
| IO-Given IO = DO | 19 | 111 | 85.38% |
| IO-Given IO > DO | 8 | 122 | 93.85% |
| IO-New IO = DO | 18 | 107 | 85.6% |
| IO-New IO > DO | 7 | 123 | 94.61% |

Figure 5: Comparison of word order rates for experimental factors of Experiment 1

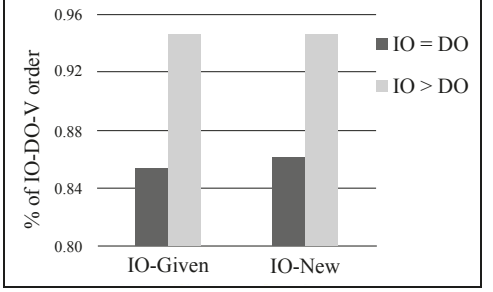



Table 6: Fixed effect values of logistic mixed-effect model for Experiment 1

| | Estimate | Std. Error | z value | Pr (>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 4.299 | 0.830 | 5.175 | < 0.0001 *** |
| Length = IO > DO | 1.479 | 0.308 | 4.802 | < 0.0001 *** |
| Givenness = IO-Given | -0.040 | 0.203 | -0.197 | 0.844 |
| IO > DO : IO-Given | -0.059 | 0.202 | -0.294 | 0.769 |
| N.B. Success corresponds to Order = IO-DO | | | | |

[14] All our tables providing raw data include only the data points that are taken into account in the analyses; erroneous/incomplete answers as well as answers presenting scarce word orders are discarded here.

[15] All models were first calculated including individual slopes for the experimental factors for items and participants. Only intercept models will be reported here since the maximal models did not converge or, when they did, did not explain significantly more variance than the corresponding intercept model.

[16] Although we did not find any effect of givenness, at this point we are not willing to conclude that the information status of the IO has no effects on the relative order between the DO and the IO. We assume that a written sentence completion task is not the best paradigm to study the effect of the information structure on word order. Hence, follow up experiments with more appropriate paradigms are necessary to study this effect.

### 3.3. Experiment 2: Bare DOs vs. Bare-modified DOs

#### 3.3.1. Method and material

In the previous experiment, we found a rather surprisingly high rate of IO-DO order for bare-modified objects. This rate is much higher than the rate observed for these DOs in the corpus data (66.7%) and even higher than the rate observed for bare (single-word) DOs in that data (84.2%).[17] This difference can be explained invoking particularities of the corpus data, such as inherent (uncontrolled) variation, possible annotation errors, or the journalistic register of the corpus. It may also be the case that the humanness of IOs is favoring the IO-DO-V order. Several studies on sentence production across different languages of the world (e.g. English, German, Greek, Spanish, and Japanese) suggest that animate constituents are likely to precede inanimate constituents (e.g. COLLINS 1995; BRANIGAN & FELEKI 1999; KEMPEN & HARBUSCH 2004; BRESNAN et al. 2007; BRANIGAN et al. 2008). Hence, the strong preference for IO-DO order observed in Experiment 1 may also be attributed to an "animate-before-inanimate" preference. In any event, the question of how bare-modified DOs differ in their ordering preferences from bare DOs needs to be addressed in a controlled experiment.

Recall that the latter only differ with respect to their length. Indeed, bare-modified DOs can be viewed as longer versions of bare DOs. Experiment 2 focuses only on this factor. In Experiment 1, bare DOs contained only a one-word adjunct and hence are only minimally construed as bare-modified. In Experiment 2 the relative length between bare and bare-modifed DOs is manipulated by adding a two-word adjunct. Furthermore, IOs are construed as inanimate locative arguments.

A preamble, three blank boxes and a list of four elements, containing a choice of a verb and of an IO, and two (formally identical) choices of a DO, present each item. Sixty native speakers of Persian completed each 10 target items similar to (9) below.

(9) *modir=e* *hotel* *sefāreš-kard* *hatman* ...
manager=EZ hotel recommended absolutely
[$_{DO1}$ *gol(=e* *orkide=ye* *sefid)* ] [$_{IO}$ *sar=e* *miz* ] [$_{V}$ *begozar-and* ]
flower=EZ orchid=EZ white on=EZ table put-3PL
[$_{DO2}$ *āš(=e* *rešte=ye* *Tabrizi)* ]
soup=EZ noodle=EZ of-Tabriz
'The manager of the hotel recommended they (should) absolutely put some (white orchid) flowers / some (Tabrizi noodle) soup on the table.'

Our experimental hypothesis is that bare-modified DOs have a less strong preference for the IO-DO order than their bare (unmodified) counterparts. Furthermore, if the humanness of the IOs in the previous experiment is indeed responsible for the high rate of the IO-DO-V order, we expect a lower overall rate of this order in this experiment.

[17] Note that even if we compare these rates excluding the data for the IO>DO condition the rate of 85.7% remains very high.

3.3.2. Results

Table 7 reports frequencies of ordering choices for Experiment 2. The preference for the IO-DO order is significantly less strong for bare-modified DOs ($\chi 2\,(1) = 17.27$, $p < 0.001$). Yet the rate of DO-IO order is much higher than expected, especially for bare-modified DOs.[18] These results suggest that animacy and/or the argument structure of the verb must play a significant role in ordering preferences between the DO and the IO.

Table 7: Raw data and percentages for Experiment 2

| DO type | DO-IO | IO-DO | % of IO-DO |
|---|---|---|---|
| Bare | 84 | 214 | 71.8% |
| Bare-modified | 147 | 151 | 50.7% |

3.4. Experiment 3: *Rā*-marked DOs vs. indefinite DOs

3.4.1. Method and material

Although a preference for the non verb-adjacent position of *rā*-marked DOs is quite uncontroversial, we still think there are good reasons to compare the behavior of *rā*-marked DOs and indefinite (non-*rā*-marked) DOs in a controlled experiment. Firstly, comparable data for all DO types is needed in order to propose a general analysis of word order in ditransitive constructions, yet none of the previous experiments includes *rā*-marked DOs and hence the only available quantitative data on ordering preferences of the latter is based on corpus data. Moreover, corpus data show an almost systematic preference for the DO-IO order for these DOs, on which relative length has no impact (for details see FAGHIRI & SAMVELIAN 2014: 226). This suggests a near categorical preference triggered by the presence of *=rā*, even though its absence does not trigger the inverse order. While, it is conceivable for *rā*-marked DOs, occupying the highest degree of determination for a DO, to display less variation than indefinite DOs placed lower on that scale, one would still expect some degree of variation comparable to bare DOs placed on the other end of the scale. Secondly, Persian, as we have mentioned before, has two ways to mark the indefinite determination on a noun phrases that hereafter we refer to as *yek*-marking, ex. *yek medād,* and *i*-marking, ex. *medād=i*. It is interesting to verify if these two alternatives behave alike with respect to word order. While the two markings are semantically equivalent,[19] they are not formally identical and with respect to the grammatical weight, it is plausible to assume that *yek*-marking is heavier than *i*-marking. The latter is an enclitic and yields one graphical and phonological word, while the former yields two graphically and phonological separate words. Given this fact, we expect to observe a higher rate of the DO-IO order for *yek*-marked DOs than for *i*-marked DOs.

[18] The mixed-effect logistic regression analysis of these data will be presented with Experiment 3 in Section 3.4.

[19] Note that these two markings do not share the same register. *Yek*-marking is used in the formal as well as the colloquial register – in which case it is phonologically simplified (*yek* > *ye*) –, while *i*-marking has a formal connotation.

This experiment follows a design with the three following conditions: 1) *Rā*-marked DO, 2) Indefinite DO with *yek*, 3) Indefinite DO with *=i*. All IOs are construed as human beneficiary arguments. As for Experiment 2, each item is presented by a preamble, three blank boxes and a list of four elements, containing a choice of a verb and an IO, and two (formally identical) choices of DOs. The experiment included 15 target items similar to (10).

(10) *ba'd_az* *šām* *sarāšpaz*
after supper head_cook
[$_{DO1}$ a. *keyk=i* b. *yek keyk* c. *keyk=rā* ] [$_{IO}$ *barā=ye mehmān-ha* ] [$_{V}$ *āvard* ]
cake=INDF a cake cake=DOM for=EZ guest-PL brought
[$_{DO2}$ a. *baste=i* b. *yek baste* c. *baste=rā* ]
package=INDF a package package=DOM
'After the dinner, the head-cook brought a/the cake/package for the guests.'

Experiment 3 is conducted via the same questionnaire as for Experiment 2 and the participants of the two experiments are the same. However, due to a technical error, we had to disregard the answers of our first twenty-one participants for this experiment.

3.4.2. Results

Table 8: Raw data and percentages for Experiment 3

| DO type | DO-IO-V | IO-DO-V | % of DO-IO |
|---|---|---|---|
| *rā*-marked | 163 | 31 | 83.60% |
| *yek*-marked | 117 | 78 | 60% |
| *i*-marked | 127 | 68 | 65.5% |

Table 8 presents the results of Experiment 3. *Rā*-marked DOs, while having a strong preference for the DO-IO order, present more variation than observed in the corpus data (83.6% vs. 96%). Consistent with our experimental hypotheses, 1) indefinite DOs – the two types taken together – display a more moderate, yet clear, preference for this order (62.75%), and 2) the DO-IO preference is slightly though not significantly stronger for *i*-marked DOs than for *yek*-marked ones ($\chi 2$ (1) = 0.89, p = 0.35).

Since Experiment 2 and Experiment 3 are conducted via the same questionnaire, we fitted a mixed-effect logistic regression model, combining the results of the two experiments. In order to have a more compact model, we have grouped *i*-marked and *yek*-marked indefinite DOs. Table 9 provides the results of this model. The estimated coefficients and the intercept (which corresponds to the value for DO=Bare), all statistically significant and hence directly comparable,[20] suggest as expected that the "default" word order[21] is IO-DO for bare DOs and DO-IO for indefinite and *rā*-marked DOs. They also highlight the gradual nature of word order preferences for different DO types, namely the fact that bare-modified DOs and indefinite DOs show a weaker preference for their respective default word order.

[20] Note that positive coefficients vote for the DO-IO order and negative coefficients for IO-DO.

[21] The estimated probability of the DO-IO-V order is $0.23_{CI: [0.12 - 0.38]}$, $0.5_{CI: [0.39 - 0.61]}$, $0.62_{CI: [0.43 - 0.78]}$ and $0.88_{CI: [0.77 - 0.94]}$, respectively for Bare, Bare-modified, Indefinite and *Rā*-marked DOs.

Table 9: Fixed effect values of logistic mixed-effect model fitted to the combination of data from Experiment 2 and 3

| | Estimate | Std. Error | z value | Pr (> \|z\|) |
|---|---|---|---|---|
| DO = Bare (Intercept) | -1.213 | 0.368 | -3.307 | < 0.0001 *** |
| DO = Bare-mod | 1.223 | 0.218 | 5.615 | < 0.0001 *** |
| DO = Indefinite | 1.711 | 0.393 | 4.349 | < 0.0001 *** |
| DO = *rā*-marked | 3.192 | 0.392 | 8.136 | < 0.0001 *** |
| N.B. Success corresponds to Order=DO-IO-V | | | | |

3.5. Experiment 4: Relative order between the Subject and the *Rā*-marked DO

3.5.1. Method and material

Our corpus data on the distribution of word order in transitive sentences revealed very little variation and interestingly showed that the degree of definiteness of the DO remains highly relevant for the relative order between the subject and the DO. Recall that only *rā*-marked DOs occur in OSV order in our dataset. In addition to obtaining experimental data on the relative order in transitive sentences, as a benchmark to compare to word order distributions in ditransitive sentences, the goal of this experiment is to study the effect of the relative length and animacy on the relative order between the subject and the DO. To this end, we focused on *rā*-marked DOs construed as human arguments to give the OSV order the highest chance.

(11) *mādarbozorg* *goft* *Āzāde* *asabāni* *ast* *čon* ...
grandmother said Azadeh angry is because
[Subj *Alirezā* / *sarosedā* ] [DO *bačče=rā*
Alireza noise child=DOM
*(ke* *yek-o-nim* *daraje* *tab* *dār-ad)* ]
that one-and-a-half degree fever have
[V1 *bidār* *karde* *ast* ] / [V2 *xaste* *karde* *ast* ]
awake done has exhausted done has
'Grandma said that Azadeh is angry because Alireza / the noise has awoken / exhausted the child (that has a temperature of 103 F).'

The experiment follows a 2x2 design. We manipulated the length of the DO, via a relative clause modifier, yielding DO-long and DO-short conditions, and the animacy of the subject, yielding Sub-animate and Sub-inanimate conditions.[22] A preamble, three blank boxes, and a list of four elements, containing a subject, a DO and two choices of verbs, constitute each item. The experiment included 20 target items similar to 0). Experiment 4 is conducted via the same questionnaire as for Experiment 2 and 3.

[22] The experiment contains one additional condition, yielded by manipulating the position of *=rā* on condition DO-Long/Sub-Anim which is not relevant for the issue at stake and hence we will not discuss in this paper. Note that this has no impact on the results reported here.

Following our observations on the effect of relative length on ordering preferences between the two objects, we expect a less important effect of length in this experiment, given that all DOs are already highly (conceptually) accessible.

3.5.2. Results

Table 11: Raw data and percentages for Experiment 4

| Conds. | SOV | OSV | % of SOV |
|---|---|---|---|
| Sub-Anim DO-Long | 229 | 11 | 95.41% |
| Sub-Anim DO-Short | 228 | 9 | 95.20% |
| Sub-Inanim DO-Long | 208 | 26 | 88.88% |
| Sub-Inanim DO-Short | 212 | 24 | 89.83% |

Figure 10: Comparison of word order rates for experimental factors of Experiment 14

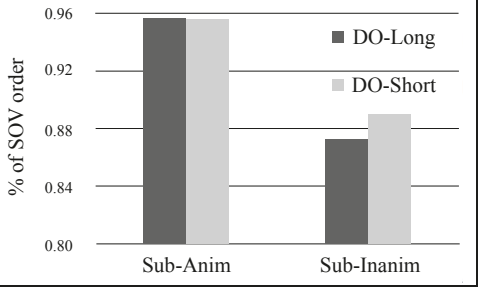



Table 11 presents the raw data, Figure 10 summarizes the distribution of word order choices with respect to our two experimental factors, and Table 12 provides the results of our mixed-effect logistic regression model. The results show an overall rate of 92.6% for the SOV order. We did not find any effects of relative length but a significant effect of animacy. The rate of OSV is higher when the subject is inanimate (10.6% vs. 4.2%).

Table 12: Fixed effect values of logistic mixed-effect model for Experiment 4

| | Estimate | Std. Error | z value | Pr (>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 3.745 | 0.325 | 11.536 | < 0.0001*** |
| Animacy = Sub-Anim | 0.400 | 0.172 | 2.329 | 0.019 * |
| Length = DO-Long | -0.101 | 0.155 | -0.654 | 0.513 |
| DO-Long: Sub-Anim | -0.012 | 0.155 | 0.078 | 0.938 |
| N.B. Success corresponds to Order = SOV | | | | |

3.6. Conclusions

The findings of our new experiments reinforce the conclusions of previous corpus and experimental studies and, in addition, shed more light on the impact of functional factors such as relative length and animacy on ordering preferences between constituents.

Our recent results confirm the following findings put forward in previous studies:

1) *rā*-marked DOs show a strong preference for the DO-IO-V order;
2) Indefinite (non-*rā*-marked) DOs show a clear, but less strong, preference for the DO-IO-V order, and an important degree of variation depending on (functional) properties of both the DO and the IO;
3) Bare single-word (i.e. non-modified) DOs have a strong preference for the IO-DO-V order;

4) Likewise, bare-modified DOs while having a general preference for the IO-DO-V order show an important degree of variation.
5) Relative length has a significant effect on word order, in ditransitive constructions, corresponding to the "long-before-short" preference.

Additionally, they allow us to make the following observations:
1) Semantic properties of the IO seem to have an impact on its relative order with respect the DO. Beneficiary/human IOs appear to be more likely to precede the DO than locative/inanimate IO.
2) In transitive constructions, humanness has a significant effect on word order corresponding to the "animate-before-inanimate" preference, but relative length does not seem to have an effect.

Most importantly, combining the results from the three last experiments, carried out via a unique questionnaire, enables us to compare between word order distribution in ditransitive and transitive constructions directly. As illustrated in Table 13, for the same sample of speakers, the rate of OSV order is below 10%, while the rate of shifted ("non-canonical") orders in ditransitive constructions can at best reach 20% – if we put aside intermediate DO types (e.g. indefinite and bare-modified DOs) or consider a non-binary criterion. Indeed, with respect to the latter, we observe that a gradual scale based on the degree of determination is much more accurate than any binary criteria. Furthermore, we observe that bareness is more accurate than markedness.

Table 13: Rates of shifted orders in transitive vs. ditransitive constructions in the data from Experiments 2, 3 and 4

| **Construction** | | **Canonical order** | | **% of shifted** |
|---|---|---|---|---|
| **Transitive** | **Observed order** | SOV | OSV | |
| | SOV | 877 | 0 | |
| | OSV | **70** | 0 | **7.39%** |
| **Ditransitive** | | DO-IO | IO-DO | |
| Bare & Marked DOs only | DO-IO | 163 | **84** | |
| | IO-DO | **31** | 214 | **23.37%** |
| All DOs: | | | | |
| Binary scale based on markedness | | 163 | **475** | |
| | | **31** | 511 | **42.88%** |
| Binary scale based on bareness | | | | |
| | | 407 | **231** | |
| | | **177** | 365 | **34.58%** |
| Gradual scale based on deg. of determination: | | | | |
| Bare=0, Bare-mod=0.25, Indef=0.75, Marked=1 | | 481.5 | **156.5** | |
| | | **79.5** | 462.5 | **20.00%** |

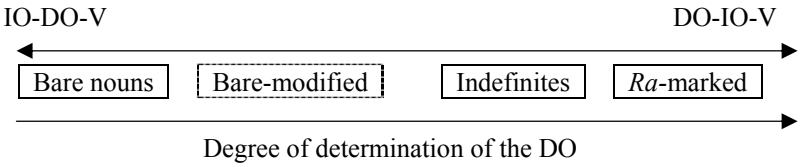
In the following section, building on these observations, we propose an account of the relative order between the DO and the IO based on functional factors and well-established cross-linguistically valid preferences.

## 4. Accounting for the relative order between DO and the IO in Persian

First of all, we will discuss the DOM criterion. The observations summarized above converge on the same simple fact that a binary distinction cannot account for the data. Ordering preferences rather follow a cline based on the degree of determination of the DO[23] as illustrated in Figure 14. On one hand, bare nouns, which occupy the lowest position on this cline, favor the IO-DO-V order and, on the other hand, *rā*-marked DOs, occupying the highest position, favor the DO-IO-V order. Everything being equal, indefinite (non-*rā*-marked) DOs occur significantly more frequently in the DO-IO-V order rather than the inverse order, highlighting that, contrary to existing hypotheses, the critical feature to determine the "default" word order is not *rā*-markedness, but bareness or absence of determination. Consequently, any dualistic account based on *rā*-markedness, and *a fortiori* an analysis of the VP in terms of dual syntactic positions such as in TOPH (cf. Scheme 6, in Section 1.2.), is incompatible with the data.

Furthermore, our data show that word order variation is much more important between the DO and the IO, than between the subject and the DO. The latter clearly converge with the SOV canonical order and are affected by functional factors much less strongly than the former. Firstly, we did not find any effects of relative length. Secondly, the degree of determination has limited impact, given that *rā*-marked DOs display some variation in our data while non-*rā*-marked DOs show none. Finally, animacy (of the subject) shows a significant but similarly limited effect – it can provoke some variations with *rā*-marked human DOs. On the contrary, all these factors are shown to have a much stronger effect on the relative order between the DO and the IO.

Figure 14: Word order cline between the DO and the IO




In the remainder of this section, we will propose an account of these word order preferences in terms of the interaction of these functional factors, or soft constraints. We will show how the different tendencies observed in our data converge into the general cross-linguistically established tendency to produce more (conceptual) accessible constituents earlier in the sentence.

The primary factor that determines the relative position of the DO with respect to IO is the degree of determination. The degree of determination reflects the degree of discourse accessibility of DOs, which is a well-known dimension of conceptual accessibility. Recall

[23] Strictly speaking, bare-modified DOs do not differ from bare DOs on their degree of determination. Nevertheless, in a larger sense and from a semantic and conceptual point of view bare-modified DOs are more determined than bare nouns. Therefore, we have decided to include them in the cline in Figure 14 only in dashed lines.

from our discussions in Section 1.1. on different realization of the DO that the cline described in Figure 14 coincides with definiteness and consequently can be mapped onto a hierarchy of discourse accessibility, such as the *(Referential) Givenness Hierarchy* proposed by GUNDEL et al. (1993). This amounts to saying that the more the DO is discourse accessible, the more it is likely to appear before the IO in the sentence. In other words, this cline converges with the well-established "given-first" preference in sentence production (cf. e.g. CLARK & HAVILAND 1977; WASOW 2002, BRESNAN et al. 2007).

Furthermore, our results show that relative length and animacy of the constituents can modify ordering preferences following the "long-before-short" and "animate-before-inanimate" tendencies. As we mentioned previously, the latter is a preference observed in many typologically different languages, accounted for in terms of the general preference in sentence production to produce more conceptually accessible constituents earlier in the sentence,[24] while the former is shown to be restricted to the preverbal domain.[25]

In line with previous studies (FAGHIRI & SAMVELIAN 2014; FAGHIRI et al. 2014), we subscribe to the analysis of this preference in terms of conceptual accessibility (cf. YAMASHITA & CHANG 2001). The account among other things builds on the following rationale: the extra material in longer constituents enhances their meaning. In other words, longer constituents are lexically and semantically richer and hence are conceptually more accessible than their shorter counterparts. Hence, the "long-before-short" preference can be accounted for in terms of a general preference to put more accessible constitutes before less accessible ones.

The three preference patterns mentioned here involve different dimensions of what is generally called conceptual accessibility and constitute different instances of the general tendency to place more conceptually accessible constituents earlier in the sentence, in short the "salient-first" preference. The more salient a DO the more it is likely to be placed before the IO in a ditransitive construction. It goes without saying that these different dimensions of conceptually accessibility do not affect the linear order in the same manner, that is, by exhorting the same strength, and interact with each other. For instance, we have seen that definiteness of the DO is a much stronger factor than relative length or animacy.[26] More importantly, they interact with another dimension of the conceptual accessibility that is the hierarchy of the grammatical roles. Note that our empirical study on the relative order between the subject and the DO is also compatible with this idea and suggest that the way different dimensions of conceptual accessibility can affect the linear order depends on the hierarchy of grammatical roles. However, more experiments are necessary to pin down the exact nature of the interaction of these factors.

---

[24] There is an ongoing debate in the literature on how exactly conceptual accessibility affects word order. Some suggests that animacy only influences word order in an indirect manner via grammatical function assignment, given KEENAN & COMRIE (1977)'s *NP Accessibility Hierarchy*. Other studies, including our work, show that animacy can directly affect the word order as well.

[25] Indeed, the mirror-image preference holds in the postverbal domain.

[26] Also, recall that our experimental data did not find a significant effect of (discourse) givenness of the IO, while its animacy seems to play a role.

## 5. Conclusions

In this paper, we addressed the issue of word order in ditransitive constructions in Persian. The existence of DOM and its implications on ordering preferences has generally led various studies to postulate a canonical word order in these constructions in Persian, comparable to those observed for the respective order of the subject and the DO. Based on corpus and experimental studies, we showed that the two phenomena are in fact not comparable. While subjects strongly prefer to precede direct and indirect objects, the order between direct and indirect objects is much more variable. Consequently, no canonical word order can be postulated for the ordering of the DO and the IO. Taken together, all corpus analyses and experiments are highly compatible with a salient-first analysis where salience can be determined by a number of factors (such as length and animacy). These results show that ordering preferences are the result of a combination of strong and weak constraints, where strong constraints may lead to ordering preferences that can be conceived as a canonical order. They are however incompatible with the assumption of equally strong constraints independent of the construction.

## References

AISSEN, Judith 2003: "Differential object marking: Iconicity vs. economy." *Natural Language and Linguistic Theory* 21(3), pp. 435-483

BONAMI, Olivier & Pollet SAMVELIAN 2015: "The diversity of inflectional periphrasis in Persian." *Journal of Linguistics* 51(2), pp. 327-382

BRESNAN, Joan, Anna CUENI, Tatiana NIKITINA & Harald BAAYEN 2007: "Predicting the dative alternation." In: Gerlof BOUMA, Irene KRÄMER & Joost ZWARTS (eds.): *Cognitive Foundations of Interpretation*. Amsterdam: KNAW, pp. 69-94

BOSSONG, Georg 1985: *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Gunter Narr

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*, pp. 96-101

BRANIGAN, Holly P., Martin J. PICKERING & Mikihiro N. TANAKA 2008: "Contributions of animacy to grammatical function assignment and word order during production." *Lingua* 118(2), pp. 172-189

BRODY, Jill 1984: "Some problems with the concept of basic word order." *Linguistics* 22(5), pp. 711-736

BROWNING, Marguerite & Ezat KARIMI 1994: "Scrambling to object position in Persian." In: Norbert COVER & Henk VAN RIEMSDIJK (eds.): *Studies on Scrambling*. Berlin: Mouton de Gruyter, pp. 61-100

CLARK, Herbert & Susan HAVILAND 1977: "Comprehension and the given-new contrast." In: Roy O. FREEDLE (ed.): *Discourse Production and Comprehension*. Hillsdale, N.J.: Ablex, pp. 1-40

COLLINS, Peter. 1995: "The indirect object construction in English: An informational approach." *Linguistics* 33(1), pp. 35-50

DABIR-MOGHADDAM, Mohammad 1992: "On the (in)dependence of syntax and pragmatics: Evidence from the postposition *-râ* in Persian." In: Dieter STEIN (ed.): *Cooperating with Written Texts: The Pragmatics and Comprehension of Written Texts*. Berlin: Mouton de Gruyter, pp. 549-573

DRYER, Matthew S. 1995: "Frequency and pragmatically unmarked word order." In: Mickey NOONAN & Pamela DOWNING (eds.): *Word Order in Discourse*. Amsterdam: John Benjamins, pp. 105-135

_____ 2007: "Word order." In: Timothy SHOPEN (ed.): *Clause Structure, Language Typology and Syntactic Description* 1. Cambridge: Cambridge University Press, pp. 61-131

FAGHIRI, Pegah 2016: *La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique*. Unpublished PhD thesis, Université Sorbonne Nouvelle Paris 3

FAGHIRI, Pegah & Pollet SAMVELIAN 2014: "Constituent ordering in Persian and the weight factor." In: Christopher PINON (ed.): *Empirical Issues in Syntax and Semantics* 10, pp. 215-232

_____ 2016: "How much structure is needed? The case of the Persian VP." In: Doug ARNOLD, Miriam BUTT, Berthold CRYSMANN, Tracy HOLLOWAY KING & Stefan MÜLLER (eds.): *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*. Stanford, CA: CSLI Publications, pp. 236-254

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*. Stanford, CA: CSLI Publications, pp. 217-237

GANJAVI, Shadi 2007: *Direct Objects in Persian* (unpublished PhD thesis, University of Southern California); https://search.proquest.com/docview/304822291

GHOMESHI, Jila 1997: "Topics in Persian VPs." *Lingua* 102(2), pp. 133-167

GUNDEL, Jeanette K., Nancy HEDBERG & Ron ZACHARSKI 1993: "Cognitive status and the form of referring expressions in discourse." *Language* 69(2), pp. 274-307

HOPPER, Paul J. & Sandra A. THOMPSON 1980: "Transitivity in grammar and discourse." *Language* 56(2), pp. 251-299

KARIMI, Simin 1999: "A Note on Parasitic Gaps and Specificity." *Linguistic Inquiry* 30, pp. 704-713

_____ 2003: "On object positions, specificity, and scrambling in Persian." In: Simin KARIMI (ed.): *Word Order and Scrambling*. Oxford: Blackwell, pp. 91-124

_____2005: *A Minimalist Approach to scrambling: Evidence from Persian*. Berlin: Mouton de Gruyter

KEENAN, Edward L. & Bernard COMRIE 1977: "Noun phrase accessibility and universal grammar." *Linguistic Inquiry* 8(1), pp. 63-99

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*. Berlin: Mouton de Gruyter, pp. 173–181

LAMBRECHT, Knud 1996: *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge : 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), pp. 177-208

MAHOOTIAN, Shahrzad 1997: *Persian.* London: Routledge

ROBERTS, John R., Behrooz BARJASTEH DELFOROOZ & Carina JAHANI 2009: *A Study of Persian Discourse Structure.* Uppsala: Acta Universitatis Upsaliensis

SAMVELIAN, Pollet 2001: "Le statut syntaxique des objets nus en persan." *Bulletin de la Société de Linguistique de Paris* 96(1), pp.349-388

_____ 2007: "A (phrasal) affix analysis of the Persian ezafe." *Journal of Linguistics* 43(3), pp. 605-645

_____ 2012: *Grammaire des prédicats complexes. Les constructions nom-verbe.* Cachan: Lavoisier

_____ 2018: "The *Ezâfe* Construction, Differential Object Marking and Complex Predicates." In: Anousha SEDIGHI & Pouneh SHABANI-JADIDI (eds.): *The Oxford Handbook of Persian Linguistics*. Oxford: Oxford University Press, pp. 226-269

SAMVELIAN, Pollet & Pegah FAGHIRI 2014: "Persian complex predicates: How compositional are they?" *Semantics - Syntax Interface* 1(1), pp. 43-75

SIEWIERSKA, Anna 1988: *Word Order Rules*. London: Routledge

SORACE, Antonella & Frank KELLER 2005: "Gradience in linguistic data." *Lingua* 115, pp. 1497-1524

WASOW, Thomas 2002: *Postverbal Behavior.* Stanford, CA: CSLI Publications

YAMASHITA, Hiroko & Franklin CHANG 2001: "'Long before short' preferences in the production of a head final language." *Cognition* 81(2), pp. B45-B55