Revisiting the *Rā*-Marked vs. Non-*Rā*-Marked Dichotomy in the Analysis of the Persian VP

Pegah Faghiri

Goals and Background

- An empirical approach to the analysis of the Persian VP.

- A series of quantitative studies (including corpus analyses and experiments) to evaluate available views by testing the validity of their predictions, in particular, in terms of word order variations, in line with studies on syntactic alternation, e.g. on HNPS in English (Wasow 1997, 2002; Bresnan et al. 2007).

Goals and Background

- The prevailing view of the Persian VP assumes two syntactic positions for DOs based on differential object marking (DOM) or *rā*-marking, ex.:

(1) a. $[_{VP}$ **DP$_{[+rā]}$** $[_{V'}$ PP V]]

    b. $[_{VP}$ $[_{V'}$ PP $[_{V'}$ **DP$_{[-rā]}$** V]]]

➢ **The Two Object Position Hypothesis** (adopted from Karimi, 2003:105)

N.B.: Most studies formulate this claim in terms of a binary feature such as specificity assumed to trigger *=rā* (cf. e.g. Karimi 2003:91).

Outline

1. The Two Object Position Hypothesis
2. Neutral/unmarked/canonical word order
3. Data
   1. Corpus analyses
   2. Follow up experiments
4. Discussion

The Two Object Position Hypothesis

- It is claimed that *rā*-marked and non-*rā*-marked DOs display several syntactic and semantic asymmetries that can be straightforwardly accounted for if two distinct syntactic positions are posited for each type.

  (e.g. Karimi 1990, 1994, 2003, 2005, Browning & E. Karimi 1994, Ghomeshi 1997, Ganjavi 2007, Modaressi 2014, notable exception: Samvelian 2001, Bonami & Samvelian 2015)

- Despite significant differences, these studies assume that (in spell out) *rā*-marked DOs appear in a higher syntactic position than their non-*rā*-marked counterparts, cf. VP external vs. VP internal in terms of Diesing (1992).

- The backbone argument put forward in support of this view relies on an broadly assumed (theoretical) hypothesis on the neutral/unmarked/canonical word order.

The Two Object Position Hypothesis

According to the broadly assumed hypothesis on canonical word order in Persian, hereafter, **the DOM criterion**:

In unmarked sentences , *rā*-marked DOs precede while non-*rā*-marked DOs follow IOs.

(2) a. (S) **DO=rā** IO V

   b. (S) IO **DO** V

(3a) [$_{DO}$ ketāb=rā] [$_{IO}$ be Mina] dād-am   vs.   (3b) [$_{IO}$ be Mina] [$_{DO}$ (yek) ketāb] dād-am

          book=RA   to M   gave-1SG   to M   a   book   gave-1SG

    'I gave the book to Mina.'   'I gave a book/some books to Mina.'

(e.g. Karimi 1994, 2003, Browning & E. Karimi 1994, Mahootian 1997, Rasekhmahand 2004)

Neutral/unmarked/canonical word order

Definition: The order in which constituents appear in least pragmatically and stylistically marked or neutral sentences (Siewierska 1988)

➢ Identifying such word order among competing word orders is not (always) straightforward. Frequency is one single criterion upon which most authors rely Roughly: the more frequent word order is the less marked one (Lambrecht 1996)

➢ In some cases intuitions are strong enough to be regarded as robust and reliable. Yet, there are cases in which the difference between available alternatives are more subtle and intuitions are less stable.

➢ The quantitative approach, provided methodological standards are respected, remains the most reliable way to identify the unmarked order.

Neutral/unmarked/canonical word order

For instance, some grammars accept both orders for indefinite non-*rā*-marked DOs:

(4) a. Yusef **ketāb=rā** be ketābxane dād
 Y   book=RA to library   gave
 'Yusef gave the book to the library.'

 b. Yusef az   ketābxāne **ketāb** gereft
 Y   from library   book took
 'Yusef took a/some book(s) from the library.'

 c. Yusef az   ketābxāne **ketāb=i**   gereft/ Yusef **ketāb=i**   az   ketābxāne gereft
 Y   from library   book=INDF took   Y   book=INDF from library   took
 'Yusef took a book from the library.'

(adopted from Givi Ahmadi & Anvari, 1995:305)

Outline

1. The Two Object Position Hypothesis
2. Neutral/unmarked/canonical word order
3. Data
   1. Corpus analyses
   2. Follow up experiments
4. Discussion

Corpus data
(for details Faghiri & Samvelian 2014, Faghiri 2016)

- Bijankhan Corpus (2,6m tokens, The *Hamshahri* daily, manually tagged for POS ; Tehran University : http://ece.ut.ac.ir/dbrg/bijankhan/)

- Semi-random sample of potentially "ditransitive" sentences and manual identification of (S)-DO-IO-V and (S)-IO-DO-V patterns -> 894 tokens

- Pilot annotation of the data:

  - The relative order between the DO and the IO (binary variable)

  - *Rā*-marking of the DO (binary variable)

  - Bareness of the DO (binary variable)

Evaluating the DOM criterion

Distribution of the relative order with respect to *rā*-marking:

	rā-marked DOs	non-*rā*-marked DOs
DO-IO-V	**403 (95%)**	167
IO-DO-V	21	**303 (64%)**
Total	424	470 894

Much more variation than expected for non-*rā*-marked DOs

Evaluating the DOM criterion:

- The rate of canonical word order (against shifted word orders) is 79%.

- N.B. In a comparable sample of transitive sentences (from the same corpus) the rate of the canonical word order (SOV) is 95%.

Evaluating the DOM criterion

Distribution of the relative order with respect to *rā*-marking and bareness:

	Rā-marked DOs	Non-*rā*-marked DO	
		Bare	Non-bare
DO-IO-V	**403 (95%)**	43	**124 (62%)**
IO-DO-V	21	**228 (84%)**	75
Total	424	271	199 894

a binary classification is clearly not adequate

Corpus data
Evaluating the DOM criterion

A more fine-grained classification of **non-*rā*-marked** DO:

- Presence of an indefinite/quantified determination (with or without adjuncts)
  -> **Indefinite** DOs, ex.   *yek/se(=ta)*   *ketāb(=e tarix)*
                        a/three(=CLF) book(=EZ history)
                        'a/three (history) book(s)'

- Absence of any determination or quantification:
  – With adjuncts   -> **Bare-modified** DOs, ex. *ketāb=e tarix* 'history book'
  – No adjuncts   -> **Bare** DOs, ex. *ketāb* 'book'

Corpus data
Evaluating the DOM criterion

Distribution of the relative order
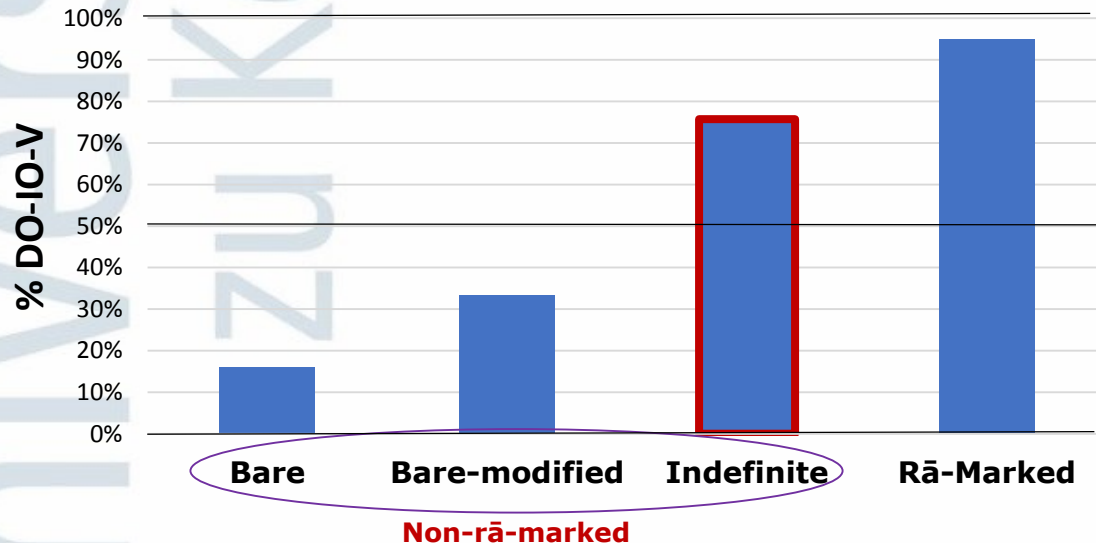
For details on the statistic analyses see Faghiri & Samvelian (2014) and Faghiri (2016:133-154)

Conform to the DOM hypothesis:
- Rā-marked DOs show a very strong preference for DO-IO-V.
- Bare (single-word) DOs show a strong preference for IO-DO-V.

In total contradiction:
- Indefinite DOs show a clear preference for DO-IO-V, grouping with rā-marked DOs rather than bare DOs.

In partial contradiction:
- Bare DOs carrying some modification while having a clear preference for IO-DO-V, show much more variation.

Experimental data
(for details Faghiri et al. 2014, Faghiri et al. 2015, Faghiri 2016)

Eliciting ordering preferences of native speakers of Persian in controlled experiments with Latin Square Design (counterbalanced and randomly ordered lists of items)
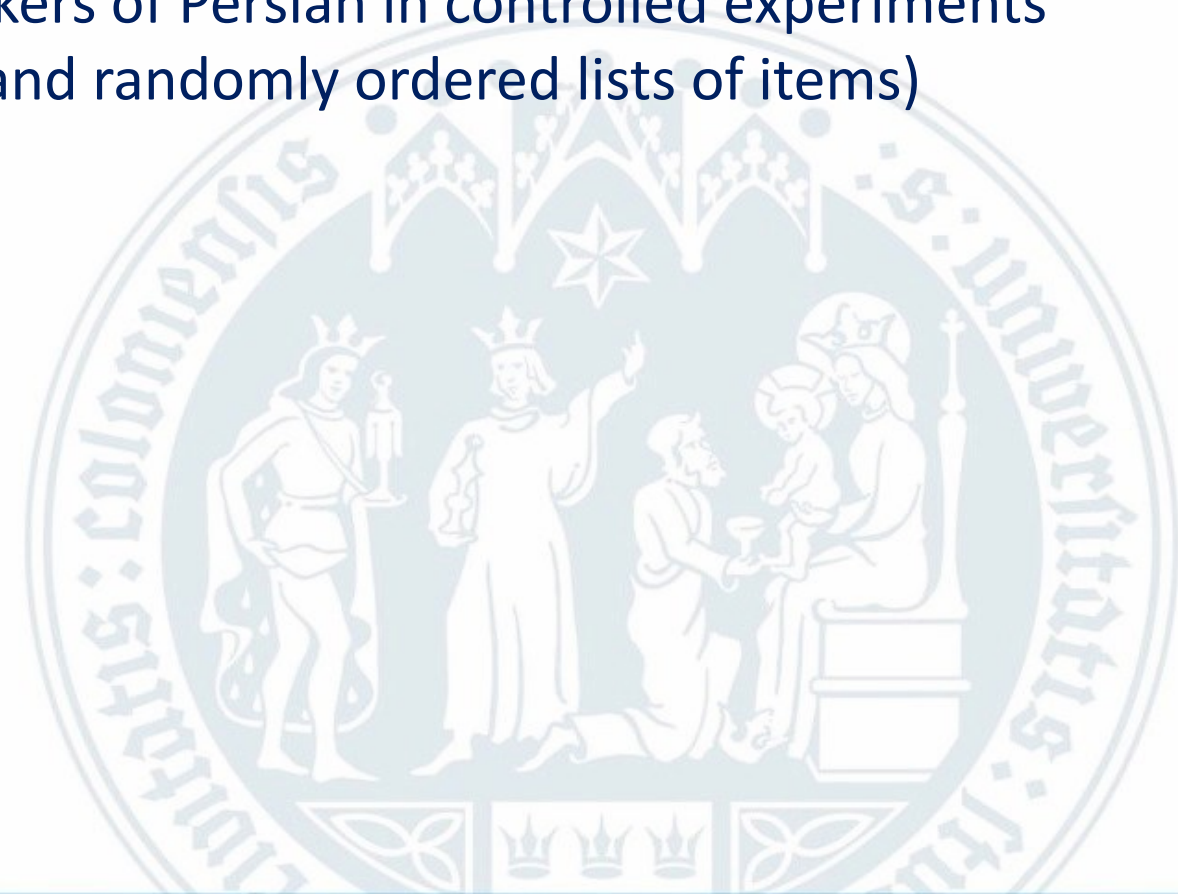
Experimental data
(for details Faghiri et al. 2014, Faghiri et al. 2015, Faghiri 2016)

➢ 4 experiments to test the relative order between DOs and IOs, for:

1. Indefinite (non-*rā*-marked) DOs (manipulating length and givenness)

2. Bare modified DOs (manipulating length and givenness)

3. Indefinite DOs (*i*-marked and *yek*-marked) vs. *Rā*-marked DOs

4. Bare DOs vs. Bare-modified DOs

(see the appendix below for more details on these experiments)

➢ 1 experiment to test the relative order between Subjects and *rā*-marked DOs
(e.g. as a "benchmark" for variation rate, *i.e.* rate of shifted vs. canonical orders)

Experimental data
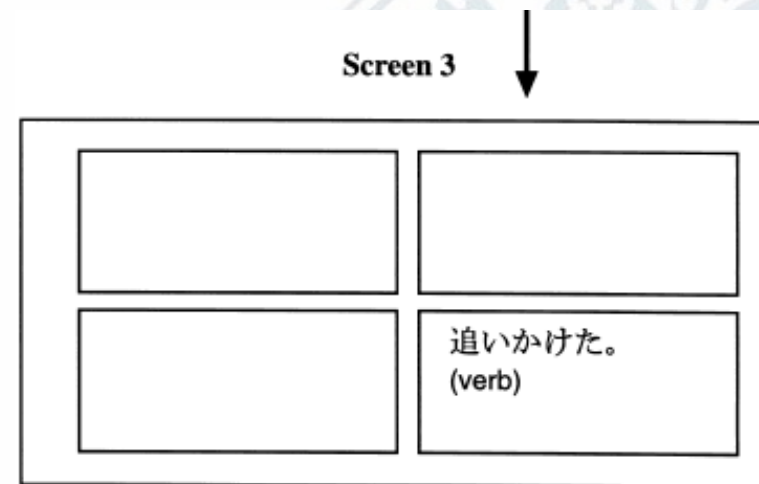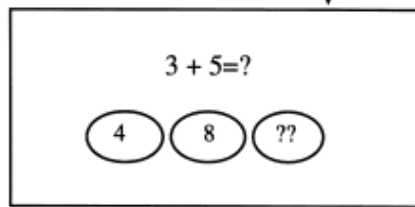# Experimental paradigm

Sentence completion task with given elements via web-based questionnaires

Inspired by <u>sentence recall task paradigms</u> (Stallings et al. 1998, Yamashita & Chang, 2001)
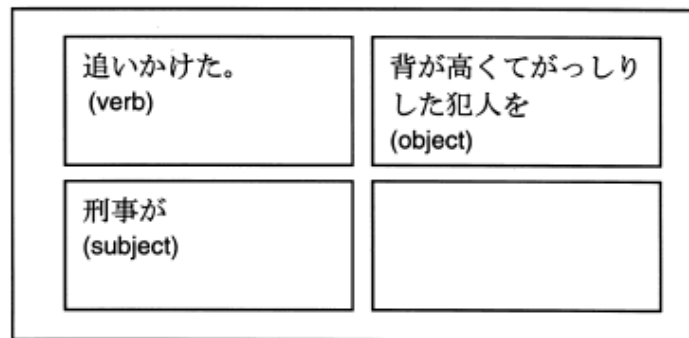
This task is said to encourage subjects to produce their sentences from the meaning

Experimental data
Example of an experimental item (Exp. 1)

کولر از کارافتاده بود و مشتری‌ها ابراز نارضایتی می‌کردند. آبدارچی وقتی اعتراضها به اوج رسید

داد.

یک لیوان شربت سکنجبین تگری

یک قاچ هنداونه‌ی بدون هسته‌ی خنک

به مشتری‌ها

ادامه ←

Example of an experimental item (Exp. 3)

دو روز بعد از آغاز کار، سخنگوی شورای شهر

.

پیشنهاد کرد
طرح جامعی
پروژهٔ جدیدی
به شهردار تهران

ادامه ←

Experimental data
Summary of results

The relative order between the DO and the IO:

- Exp. 1: Indefinite (non-*rā*-marked) DOs: DO-IO-V (mean rate: 68%)

- Exp. 2: Bare modified DOs: IO-DO-V (mean rate: 90%)

    N.B.: There is a significant (p<0.001) "long-before-short" preference
- Exp. 3: Preference for DO-IO-V: *Rā*-marked DOs > Indefinite DOs (84% vs. 63%)
- Exp. 4: Preference for IO-DO-V: Bare DOs > Bare-modified DOs (72% vs. 51%)

The relative order between the Subject and the DO: SOV (mean rate: 92%)

    N.B.: There is a significant (p<0.001) "animate-before-inanimate" preference

➢ Overall, these results are in line with our corpus findings and likewise invalidate the DOM hypothesis.

Ordering preferences of native speakers

This empirical investigation shows that ordering preferences between the DO and the IO, rather than being dichotomous on the basis of $r\bar{a}$-marking, follow a cline based on the degree of determination of the DO (roughly discourse-givenness, cf. e.g. Referential Givenness Hierarchy, Gundel et al. 1993).

IO-DO-V   **DO-IO-V**

Bare	**Bare-mod**	**Indefinite**	***Rā*-marked**
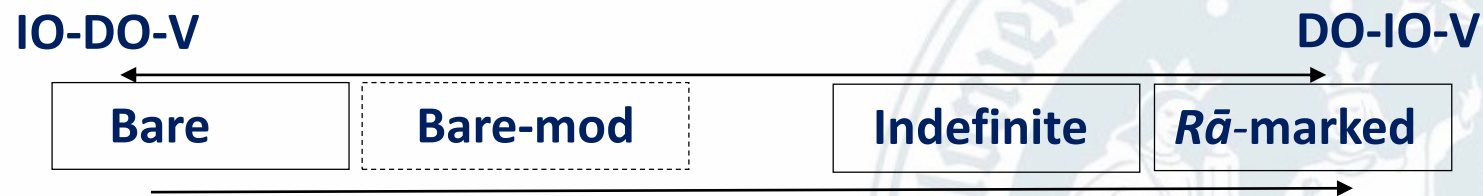
Furthermore, different ordering preferences observed in our data can be accounted for in terms of the interaction of different functional factors or soft constraints (e.g. degree of determination, length, animacy), corresponding to the general cross-linguistically established tendency to produce more (conceptual) accessible constituents earlier in the sentence.

Concluding remarks

- Word order in ditransitive constructions is not a matter of grammatical/strong constraints (*i.e.* syntactic position) but a matter of preference (soft constraints).

- These data undermine an (any) analysis of the VP in terms of a dual syntactic position based on *rā*-marking, while they supports a flat structure view of the VP, in which the linear position of verbal complements is not constrained by syntactic rules.

N.B.: For a discussion of other available arguments in favor of the TOPH (*i.e.* other asymmetries with respect to e.g. semantic (in)dependence from the verb, binding and scope relations, parasitic gaps and coordination) see Faghiri & Samvelian (2016) and Faghiri (2016:220-256).

References

Browning, M. & E. Karimi. 1994. Scrambling to object position in Persian. In N. Cover, H. van Riemsdijk (eds), Studies on scrambling, 61-100.

Diesing, M. 1992. Indefinites. Linguistic Inquiry Monographs. Cambridge, Mass: MIT Press.

Faghiri, P. 2016. La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique, PhD Dissertation, Université Sorbonne Nouvelle Paris 3.

Faghiri, P. & P. Samvelian. 2014. Constituent ordering in Persian and the weight factor. In Christopher Pinon (ed.), Empirical issues in syntax and semantics 10, 215 – 232, CNRS.

Faghiri P. & P. Samvelian. 2016, How much structure is needed? The case of the Persian VP, In D. Arnold, M. Butt, B. Crysmann, T. Holloway King & S. Müller (eds.) Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, 236–254, Stanford, CA: CSLI Publications.

Faghiri, P., P. Samvelian & B. Hemforth. 2014. Accessibility and Word Order: The Case of Ditransitive Constructions in Persian. In S. Müller (ed.): Proceedings of the 21st HPSG Conference, 217–237. CSLI Publications.

Faghiri, P. , B. Hemforth & P. Samvelian. 2015. Length effects in an OV language with Differential Object Marking and mixed head direction, Poster presented at the CUNY Sentence Processing Conference, University of Florida.

Ganjavi, S. 2007. Direct objects in Persian, Doctoral dissertation, University of Southern California.

Givi Ahmadi H. & H. Anvari. 1995. Dastur zabāne fārsi [The grammar of Persian]. Mo'assese farhangi Fātemi.

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

Gundel, J. K., N. Hedberg & R. Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. Language 69(2). 274–307.

Karimi, S. 2003. Object positions, specificity and scrambling. In Karimi, S. (ed.) Word Order and Scrambling, 91-125.

Mahootian, S. 1997. Persian, New York: Routledge.

Modarresi, F. 2014. Bare nouns in Persian: Interpretation, grammar and prosody, PhD dissertation, Humboldt-Universität zu Berlin.

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

Siewierska, A. 1988. Word order rules. Routledge.

Stallings, L. M., M. C. MacDonald & P. G. O'Seaghdha.1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy NP shift. Journal of Memory and Language , 39(3), 392-417.

Yamashita, H. & F. Chang. 2001. 'Long before short' preferences in the production of a head final language, Cognition 81(2): B45-B55.

…

Acknowledgements

Pollet Samvelian

Université Sorbonne Nouvelle - Paris 3

UMR « Mondes iranien et indien »

Barbara Hemforth

CNRS – Université Paris Diderot - Paris 7

UMR « Laboratoire de linguistique formelle »

Labex « Empirical Foundations of Linguistics » (EFL)

Exp. 1: Indefinite (non-*rā*-marked) DOs

Participants are asked to complete a preamble with two choices of formally identical DOs and one choice of an IO.

2x2 design:

Relative length (in number of words) with two conditions:

DO > IO (by adding adjectival modifiers to the DO)

DO < IO (by adding a relative clause modifier to the IO)

Givenness of IO (N.B.: DO is new by definition):

IO-Given (by mentioning its referent in the preamble)

IO-New

With control for animacy:

DO [-animate/+concrete: theme] and IO [+human: beneficiary]

7 ditransitive simplex verbs: 20 target items (combined with 30 fillers)

Exp. 1: Indefinite (non-*rā*-marked) DOs

Example of (English equivalent of) an item:

The air-conditioner had stopped working and (the clients were complaining/ it kept getting warmer). When (protests/the heat) reached a peak, the janitor …

a glass of (icy mint) syrup

a slice of (seedless fresh) watermelon

to the clients (who were frustrated from the heat)

gave [   ] [   ].

Exp. 1: Example of an item on the screen

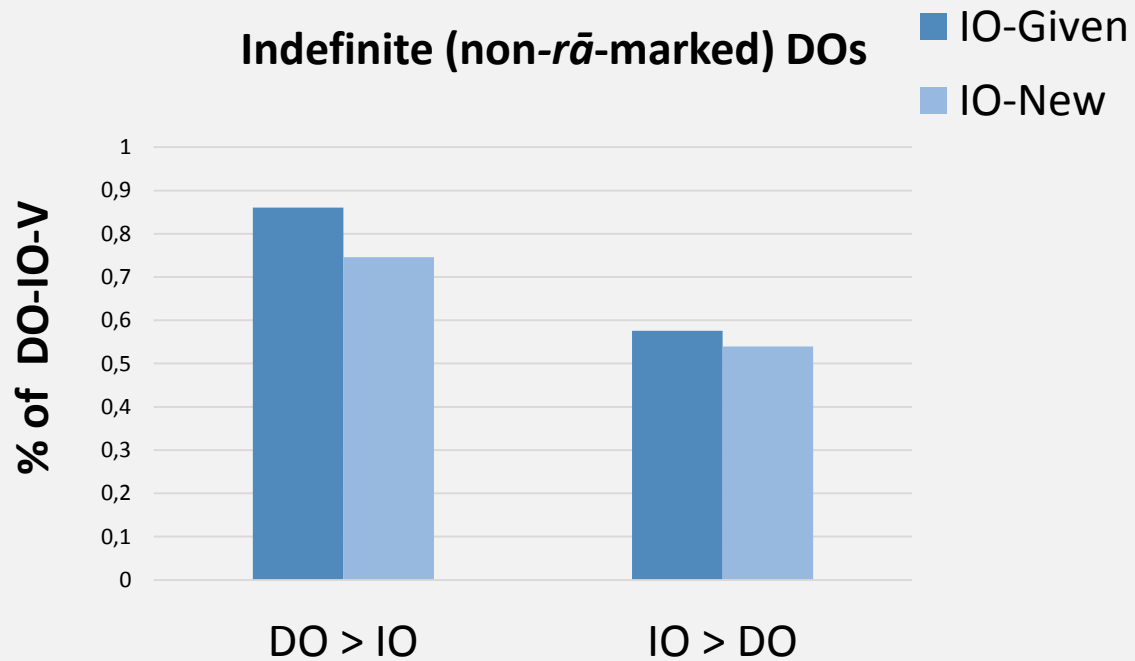کولر از کارافتاده بود و مشتری‌ها ابراز نارضایتی می‌کردند. آبدارچی وقتی اعتراض‌ها به اوج رسید

داد.

یک لیوان شربت سکنجبین تگری

یک قاچ هنداونه‌ی بدون هسته‌ی خنک

به مشتری‌ها

ادامه ←

Exp. 1: Results

Indefinite (non-*rā*-marked) DOs

Legend: ■ IO-Given ■ IO-New

Y-axis: % of DO-IO-V (0 to 1)
X-axis categories: DO > IO, IO > DO

Analysis of the data via mixed-effect regression modeling (with experimental factors as fixed effects and participant, item and verbal lemma as random effects):

- DO-IO as the default order (68% overall)

- Relative length, corresponding to a "long-before-short" preference (p<0.001)

Exp. 2: Bare-modified DOs (same design as Exp. 1)

Participants are asked to complete a preamble with two choices of formally identical DOs and one choice of an IO.

2x2 design:

      Relative length (number of words) with two conditions

            DO = IO

            DO < IO (by adding a relative clause modifier to the IO)

      Givenness of IO (N.B.: DO is new by definition):

            IO-Given (by mentioning its referent in the preamble)

            IO-New

      With control for animacy:

      DO [-animate/+concrete: theme] and IO [+human: beneficiary]

7 ditransitive simplex verbs: 20 target items (combined with 30 fillers)

Exp. 2: Bare-modified DOs

Example of (English equivalent of) an item:

The air-conditioner had stopped working and (the clients were complaining/ it kept getting warmer). When (protests/the heat) reached a peak, the janitor …

to the clients (who were frustrated from the heat)

mint syrup

cherry syrup

gave [   ] [   ].

Exp. 2: Example of an item on the screen

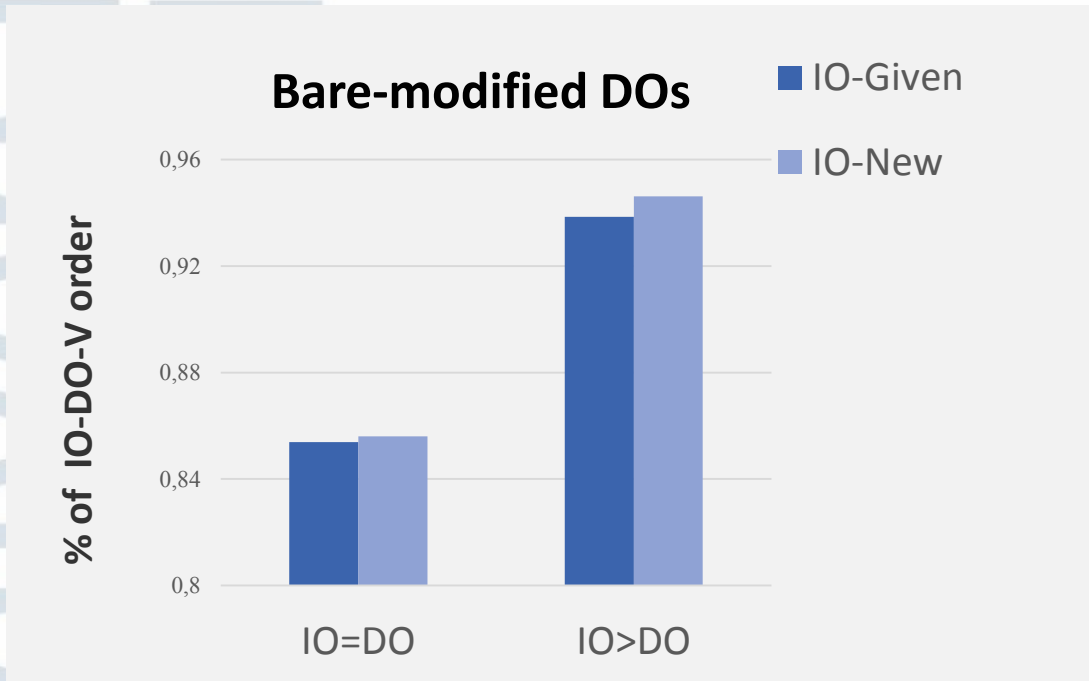کولر از کارافتاده بود و مشتری‌ها ابراز نارضایتی می‌کردند. آبدارچی وقتی اعتراض‌ها به اوج رسید

داد.

به مشتری‌ها که از گرما کلافه بودند

شربت به‌لیمو

شربت آلبالو

ادامه ←

Exp. 2: Results

Data analysis via the same modeling:

- IO-DO as the default order (90% overall)
- Relative length, corresponding to a "long-before-short" preference (p<0.001)

Exp. 3: Indefinite DOs vs. *Rā*-marked DOs

Participants are asked to complete a preamble with two choices of formally identical DOs, one choice of an IO and one choice of a verb.

Example of (English equivalent of) an item:

After the dinner, the chef ….

brought

a/the cake

a/the package

for the guests

[   ][   ] [   ].

Exp. 3: Indefinite DOs vs. *Rā*-marked DOs

Design:
   DO type with three conditions:

   1) *Rā*-marked DO : ex. *keyk=rā*

   2) Indefinite DO with *yek* : ex. *yek keyk*

   3) Indefinite DO with *=i* : ex. *keyk=i*

   With control for animacy:
         DO [-animate: theme] and IO [+human: beneficiary]

➢15 items combined with 40 fillers.

N.B.: Exp. 3 and 4 are conducted via the same questionnaire (*i.e.* same participants)

Results:

% of DO-IO-V
83.6%
60%
65.5%

Exp. 4: Bare vs. Bare-modified DOs

Participants are asked to complete a preamble with two choices of formally identical DOs, one choice of an IO and one choice of a verb.

Example of (English equivalent of) an item:

The manager of the hotel recommended they (should) absolutely ….

put

(Tabrizi noodle) soup

(white orchid) flowers

on the table

[   ][   ] [   ].

Exp. 4: Bare vs. Bare-modified DOs

Design:

    DO type with two conditions:

    1) Bare DO

    2) Bare-modified DO

With control for animacy:
        DO [-animate: theme] and IO [- animate: locative]

➢10 items combined with 45 fillers.

N.B.: Exp. 3 and 4 are conducted via the same questionnaire (*i.e.* same participants)

Results:

% of IO-DO-V
71.8%
50.7%