



Modarresi, Fereshteh and Manfred Krifka. 2021. Pseudo Incorporation and Anaphoricity: Evidence from Persian. *Glossa: a journal of general linguistics* 6(1): 124, pp. 1–32. DOI: <https://doi.org/10.16995/glossa.5875>



Open Library of Humanities

Pseudo Incorporation and Anaphoricity: Evidence from Persian

Fereshteh Modarresi, Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Schützenstr. 18, 10117 Berlin, DE
modarresi@leibniz-zas.de

Manfred Krifka, Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Schützenstr. 18, 10117 Berlin, DE

The paper investigates the discourse effects of bare nominal pseudo-incorporated objects in Persian. Contrary to claims that such nominals cannot be antecedents to anaphora, experimental results show that their anaphoric potential is only somewhat reduced, compared to indefinites. This is consonant with Krifka & Modarresi (2016), according to which they are interpreted like functional definites that are dependent on the event denoted by the verb that undergo existential closure.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1 Introduction

Many languages afford distinct realizations of a syntactic object. Massam (2001) has called one such type “pseudo incorporation” (PIN), distinguishing it from regular non-incorporated objects and from incorporated objects forming a morphological word with the predicate. PIN is related to “differential object marking” (Bossong 1985) and “semi-transitives” (Margetts 2008). It occurs in genetically and typologically diverse languages such as Hungarian (Farkas & Swart 2003), Turkish (Öztürk 2004; Kamali 2015; von Heusinger & Kornfilt 2005) and Hindi (Dayal 2011). See Massam (2009), Dayal & Sag (2019) and Borik & Gehrke (2015) for overviews.

This article discusses bare nouns (BN) in object position in Persian as an instance of PIN, contrasting them with objects with the indefinite article *yek* and with the marker *-ra*. Section 2 establishes the definitional properties of BN objects, Section 3 gives a short overview of the theoretical literature on PIN, and Section 4 develops the theory proposed in Krifka & Modarresi (2016) concerning the anaphoric properties of BNs. Section 5 reports on the results of five experiments targeting the anaphoric potential of BN objects vs. *yek*-marked objects in Persian. Section 6 discusses the theoretical implications.

Krifka & Modarresi (2016) argue that BNs in Persian are generally, and against initial appearance, interpreted like *definites* – in particular, functional definites within an *existential closure* that are *dependent* on an event variable. Three properties of PIN follow: narrow scope, number neutrality, and syntactic/prosodic integration. This analysis of BNs in Persian is in line with the similarity to *weak definites* in languages with definite articles, like English (cf. Schwarz 2013). It also allows for an analysis of BN vs. RA-marked objects in which their interpretative difference is a consequence of their syntactic position. It furthermore predicts that nouns with indefinite article *yek* are directly accessible by anaphora, whereas BNs are only indirectly accessible, hence have reduced anaphoric potential.

Our contribution relates to the last prediction, showing that BNs make natural antecedents, contrary to previous claims. It also shows that *yek*-antecedents are more easily picked up by anaphora, contrary to theories that give a similar status to these two antecedent types. And it rules out that this uptake is restricted to associative anaphora or number-neutral null anaphora.

2 Pseudo-Incorporation in Persian

The three direct object constructions contrasted in this article are bare noun (BN) objects, *yek*-marked (YK) objects, and objects with the postposition *rā* (RA) illustrated in (1).

- (1) *Maryam* (a) *ketāb* . (b) *yek ketāb* (c) *ketāb-rā* *kharid*.
 Maryam book one book book-OM bought.
 ‘Maryam bought (a) a book / books / (b) a book / (c) the book’

The BN object in (1)(a) has an indefinite, number-neutral interpretation, typical for PIN objects. The YK object in (b) has an indefinite singular interpretation, and the RA-marked object in (c) has a definite interpretation.

As evident from examples (1), Persian is an SOV language. The indefinite article, often realized as *ye*, is derived from the number word ‘one’. There is no definite article; notice that *rā* is restricted to objects and can cooccur with the indefinite marker *yek*, see below in (2)(c). The variable presence of the object marker is shared with other languages in the region, like Turkish and Hindi-Urdu.

Persian BN objects have other characteristic properties of PIN objects (cf. Massam 2009), reduced morphosyntactic expression and narrow scope with respect to other scope-bearing expressions, cf. (2a) vs. (b,c):

- (2) a. *Hæmeh ketāb khoondand.*
 all book read.PL
 ‘All people read a book/books.’ $\forall > \exists$
- b. *Hæmeh yek ketāb khoondand.*
 i) ‘All people read a book.’ $\forall > \exists$
 ii) ‘There is a book that all people read.’ $\exists > \forall$
- c. *Hæmeh yek ketāb-rā khoondand.*
 ‘There is a book that all people read.’ $\exists > \forall$

BN objects are not morphologically incorporated, i.e. do not form a word with their predicate, as they can occur with restrictive modifiers (cf. Modarresi 2014):

- (3) *Ali ketab-e-kohne kharid*
 Ali book-EZ-old bought
 ‘Ali bought old books.’

The adjective combines with the noun via the *ezafe* marker. In (3) restriction is to a subtype of books; for more specific restrictions, reference to a particular book is implicated, and the use of a RA-marked object is preferred.

Another property of PIN objects is that they are prosodically integrated with the verb. We observe prosodic integration with BN objects and also with YK objects, in contrast to RA-marked objects. Prosodic integration is indicated by focus projection in a neutral context, that is, after a question like ‘What happened?’ (Gussenhoven 1983; Selkirk 1984; Jacobs 1991) In Persian, accent is realized on the left edge of a prosodic domain.

- (4) A: *chi shodeh?* B: a. (*KETĀB khæridam*).
 what happened b. (*YEK KETĀB khæridam*).
 c. (*KETĀB-rā*) (*KHÆRIDAM*)

Prosodic separation of RA-marked objects was observed by (Hincha 1961; Browning & E. Karimi 1994; Karimi 2003) who assumed that such objects are scrambled.¹ We find a similar behavior for BN subjects: Non-integrated subjects are interpreted like singular definites, whereas integrated subjects are interpreted like number-neutral indefinites (cf. Kahnemuyipour 2009 for unaccusative verbs, Modarresi 2014 for BN subjects with a broader range of verbs as exemplified below).

- (5) a. (*KETĀB*) (*OFTAD*) b. (*KETĀB oftad*).
 book fell.3SG book fell.3SG
 ‘The book fell.’ A book / books fell.’
- (6) a. (*MOOSH*) ((*GHAZA-RĀ*) (*KHORD-E*)).
 mouse food-OM eat-PERF.3SG
 ‘The mouse ate the food’
- b. (*GHAZA-RĀ*) (*MOOSH khord-e*).
 food-OM mouse ate.3SG
 ‘The food, some mouse / mice ate.’

BN objects may also occur as complements of prepositions, such as *ketāb-rā ru ghafaseh gozāshtam* ‘I put the book on the shelf’, with no reference to a particular shelf, similar to a weak definite interpretation of *the shelf* in the English gloss (cf. Ghomeishi 2008).

PIN can be confined to certain verbs or noun-verb combinations (e.g. Asudeh & Mikkelson 2000 for Danish, Espinal & McNally 2011 for Catalan). In Persian, BN objects are quite unrestricted, e.g. *khæridan* ‘to buy’ can be combined with any BN object. But with certain verbs, like *didan* ‘to see’, *bastan* ‘to close’ and *boosidan* ‘to kiss’, BN objects do not occur easily. Causative verbs and verbs that select for objects referring to kinds require *rā*-marked objects in general (Modarresi 2010; 2014).

The property of PIN objects most relevant for the present article is their absent or reduced anaphoric potential, i.e. whether they introduce discourse referents (DRs) (cf. Farkas & de Swart 2003; Yanovich 2008; Dayal 2011; Espinal & McNally 2011; Modarresi 2014; Kamali 2015; Krifka & Modarresi 2016). For Persian, (Barjasteh 1983; Karimi 2003; Ganjavi 2007; Nemati 2010; Megerdooomian 2012) claim that BN objects cannot be antecedents of anaphoric elements. We will show that anaphoric reference to BN objects is possible but reduced in comparison to YK objects.

¹ RA marking is not a general feature of scrambling, but all RA-marked constituents (cf. Karimi & Smith 2020) appear to be scrambled.

BN objects have to be distinguished from the complements of complex predicates, a highly productive phenomenon in Persian (cf. Goldberg 1996; Samvelian & Faghiri 2014). Complex predicates generally have an idiomatized meaning. They can be transparent, such as *āb dadan* ‘water give’, ‘to water (e.g., flowers)’, or intransparent, such as *chaneh zadan* lit. ‘to chin-hit’, ‘to bargain’. Combinations of BN objects with certain verbs may develop into complex predicates, and the differentiation is not always clear-cut.

There are a number of analyses of the semantic properties of BN in Persian. Windfuhr (1979) considers them as part of the verbal predicate, a notion that does not distinguish them from the complement of complex predicates. Ghomeshi (2003) and Karimi (2003) analyze them as non-specific and non-referential, but this cannot be a defining property in the sense that all non-referential objects have to be expressed as BN:

- (7) *Leila mikhad ye mashin be-khareh.*
 Leila want.3SG a car SUBJUNCTIVE-buy.3SG
 ‘Leila wants to buy a car.’

Ghomeshi (2008) distinguishes between BNs, analyzed as NPs, vs. other kinds of objects, analyzed as DPs or QPs, but it is unclear which semantic effect this distinction has.

The distinction between BN objects with and without *rā*, cf. (1a) vs. (1c), is similar to the one between weak definites and strong definites in languages with definite article (cf. Poesio 1994; Carlson & Sussman 2005; Carlson et al. 2013; Schwarz 2014). The weak definite interpretation is illustrated in (8); Max and Mary may have read more than one, and different newspapers. This is similar to the Persian BN object example in (9).

- (8) *Max read the newspaper and Mary did, too.*
 (9) *Max rooznameh khoond. Maryam ham hamin-tor.*
 Max newspaper read.3SG Maryam also same-way

However, weak definites tend to be restricted to predicate-argument combinations that are “nameworthy”, resulting in an enriched meaning, cf *go to the hospital* (in order to get medical treatment) vs. *go to the arena*. This difference might be due to the fact that BN objects are marked differently from regular objects, a distinction that is lacking with weak definites, which may necessitate that they are licensed by the existence of a conventionalized activity.

3 Theories for the interpretation of PIN

Several theories target the properties of PIN objects and weak definites. We illustrate these theories with (1a), *ketāb kharid* ‘bought a book/books’; our focus will be on the predictions concerning anaphoric uptake.

According to Carlson (1977), bare plurals in English refer to kinds. This proposal appeared attractive for BNs in Persian, see Ghomeishi (2008), inspired by Hinch (1961) and Megardoomian (2012); see Dayal (2003; 2011) for Hindi, and Aguilar-Guevara & Zwarts (2010) for weak definites in English and German. This is illustrated in (10), where z represents the subject argument, disregarded here, *liber* is a name for the kind of books, and R is a relation that relates kinds to specimens. This representation is taken to predict that uptake of y is not possible; however, this depends on the nature of the existential quantifier $\exists y$; as a dynamic quantifier, it would allow for anaphoric uptake.

$$(10) \quad \lambda x \exists y [R(x, y) \wedge \text{bought}(y)(z)](\text{liber}) \\ = \exists y [R(\text{liber}, y) \wedge \text{bought}(y)(z)]$$

Notice that with kind-referring predicates, objects require RA marking, which is a problem for the kind analysis of BN objects (cf. Modarresi 2010; 2014):

$$(11) \quad \text{Razi alcohol-r\bar{a} kashf-kard.} \\ \text{Razi alcohol-OM discover-did.3SG} \\ \text{'Razi discovered alcohol.'}$$

Cf. also Schwarz (2014) and Espinal & Cyrino (2017) for a critical discussion of the kind-referring analysis of weak definites.

(Van Geenhoven 1992; McNally 1995; Cohen & Ertheshik-Shir 2002; Dayal 2011; Espinal & McNally 2011) argued that incorporated objects are properties that involve the existence of an entity, resulting in an analysis similar to (12). This predicts that anaphoric uptake to the entity is impossible. But van Geenhoven (1998) allows for dynamic existential quantifiers, predicting that incorporated objects support anaphoric uptake, just as regular indefinites would do.²

$$(12) \quad \lambda P \exists x [P(x) \wedge \text{bought}(x)(z)](\lambda x [\text{book}(x)]) \\ = \exists x [\text{book}(x) \wedge \text{bought}(x)(z)]$$

Chung & Ladusaw (2004) discuss a combination of predicate with an argument beyond binding, by restriction and saturation, cf. (13). Again, it depends on the nature of the existential quantifier whether anaphoric uptake is possible.

$$(13) \quad \text{RESTRICT}(\text{book}, \lambda x [\text{bought}(x)(z)]) = \lambda x [\text{book}(x) \wedge \text{bought}(x)(z)] \\ \text{SATURATE}(\lambda x [\text{book}(x) \wedge \text{bought}(x)(z)]) = \exists x [\text{book}(x) \wedge \text{bought}(x)(z)]$$

Asudeh & Mikkelsen (2000) propose that PIN objects do not introduce DRs but may allow for “inferential pronominalization”, i.e. bridging or associative anaphora (cf. also Schwarz 2019 for

² Schwarz (2013) assumes in addition the construction of event kinds, which makes dynamic quantification impossible.

weak definites). This predicts a preference for anaphoric uptake by full definite DPs, as in *John was driving down the street. The steering wheel was cold.*

Farkas & de Swart (2003) propose that PIN objects in Hungarian do not introduce DRs but that DRs may be accommodated, specifically by null anaphora for some speakers. However, it is unclear why overt anaphora cannot achieve this; if the DRs are accommodated, we should also expect a preference for definite descriptions. Also, there are technical problems with the implementation (cf. Yanovich 2008; Krifka & Modarresi 2016).

Modarresi (2015) proposes that BN objects in Persian introduce number-neutral DRs, predicting that null anaphora, lacking a number specification, can pick them up more easily. Also, if world knowledge suggests a singular or plural interpretation of the number-neutral DR,³ overt singular or plural anaphora should be acceptable. This theory predicts that BN objects support anaphoric uptake.

Concerning anaphoric uptake, we derive the following hypotheses from the literature:

- (14) BN-0: BN objects do not allow for uptake (many authors)
 BN=YK: BN objects allow for uptake to the same degree as YK objects
 (van Geenhoven 1998 for PIN, Modarresi 2014; no reasons that DRs introduction by dynamic \exists quantifier or number-neutral DRs differ from regular of DRs)
 BN-DD: BN objects allows for anaphoric uptake, preferring definite descriptions
 (Aduseh & Mikkelsen 2000, Schwarz 2019)
 BN-NL!: BN objects only allow for uptake by null anaphora
 (Farkas & de Swart 2003)
 BN-NL: BN objects prefer uptake by null anaphora (Modarresi 2014; 2015)

The theory of Krifka & Modarresi (2016) will be discussed in detail in Section 4. Rejection of hypothesis BN-NL! would be consistent with this theory; additional hypotheses will be presented in Section 4.5, cf. (35) below.

4 PIN as functional expressions under existential closure

4.1 Syntactic structure

We assume that BN objects occur at their base position within the VP whereas *rā* objects undergo scrambling (cf. Browning & E. Karimi 1993, Karimi 2003), cf. (15a,b). Subjects are typically moved to the specifier of TP but may stay within vP, cf. (c).

³ For example, *khooneh kharidan* ‘to buy house’ is biased towards a singular interpretation, whereas *havij kharidan* ‘to buy carrot’ is biased towards a plural interpretation.

- (15) a. $[_{TP} Sarah_1 [_{vP} ketāb-rā \exists [_{vP} t_1 [_{vP} t_2 [_{v^e} kharid]]]]]]$ ‘Sarah bought the book’
 b. $[_{TP} Sarah_1 \exists [_{vP} t_1 [_{vP} ketāb [_{v^e} kharid]]]]$ ‘Sarah bought a book/books’
 c. $[_{TP} [_{vP} ghaza-rā_2 \exists [_{vP} sag [_{vP} t_1 khord]]]]$ ‘The food, the dog ate’

The non-extended vP forms a maximal prosodic domain, predicting accent on *kharid* in (15)(a), *ketāb* in (b) and *sag* in (c) (cf. (4c), (4b) and (6b), Modarresi 2014). We follow Diesing (1992) in assuming that there is an existential closure over the extended verbal predicate, more specifically over the smallest vP, indicated by “ \exists ” in (15).

Dayal (2011) argues against an analysis of PIN objects as non-scrambled, in-situ objects. However, her motivating examples all involve focus movement, which is different from scrambling (cf. Frey 2010): Focus-moved expressions are interpreted in their in-situ position. In Persian, focus movement does not require RA marking, and the PIN properties like number neutrality and narrow scope, are retained.⁴

- (16) $[_{ForceP} KETĀB_2 [_{TP} hamē [_{vP} diruz [_{vP} t_1 [_{vP} t_2 [_{v^e} kharidand]]]]]]]$
 book everybody yesterday bought.3PL
 ‘It is a book / books that everybody bought yesterday.’

4.2 Interpretation in DRT

We will assume Discourse Representation theory (DRT) as in Kamp & Reyle (1993). In DRT, sentences and discourses are interpreted as discourse representation structures (DRSs), pairs of a set of DRs and a set of formulas, represented in boxes. For example, the arguments in (17) introduce the DRs x_1 , x_2 anchored to Maryam and a book of cardinality 1, and tense introduces the DR e_1 anchored to an event of x_1 picking up x_2 , cf. (20)(a). We ignore the contribution of past tense.

- (17) *Maryam yek ketāb-rā bardasht*
 Maryam one book-OM picked.3SG
 ‘Maryam picked up a book’

Sentence (18) takes (20)(a) as input, generating a new DRS (20)(b) in which the DR x_2 is picked up, and a new DRs x_3 and e_2 are introduced.

- (18) *va be yek doost-i dād*
 and to one friend-DAT gave.3SG
 ‘and gave it to a friend.’

⁴ This includes cases of contrastive topic such, as in a case pointed out by a reviewer: *Ketāb, man HAR RUZ mi-khun-am* ‘book I every day DUR-read-1SG’, ‘Books I read every day’.

Sentence (19) introduces e_3 and picks up x_1 and x_2 via null anaphora for the subject, as expressed by singular agreement, and the object; cf. the DRS (20)(c), where the contribution of *foran* ‘immediately’ remains unexpressed

- (19) *Oo ham foran khoond-esh.*
 (s)he also immediately read-did.3SG
 ‘He read it immediately.’

(20)

x_1 x_2 e_1
$x_1 = \text{Maryam}$ $\text{book}(x_2)$ $ x_2 = 1$ $e_1: \text{pick}(x_1, x_2)$

(a) DRS after (17)

x_1 x_2 e_1 x_3 e_2
$x_1 = \text{Maryam}$ $\text{book}(x_2)$ $ x_2 = 1$ $e_1: \text{pick}(x_1, x_2)$ $\text{friend-of}(x_1, x_3)$ $ x_3 = 1$ $e_2: \text{give}(x_1, x_2, x_3)$

(b) DRS after (18)

x_1 x_2 e_1 x_3 e_2 e_3
$x_1 = \text{Maryam}$ $\text{book}(x_2)$ $ x_2 = 1$ $e_1: \text{pick}(x_1, x_2)$ $\text{friend-of}(x_1, x_3)$ $ x_3 = 1$ $e_2: \text{give}(x_1, x_2, x_3)$ $e_3: \text{read}(x_1, x_3)$

(c) DRS after (19)

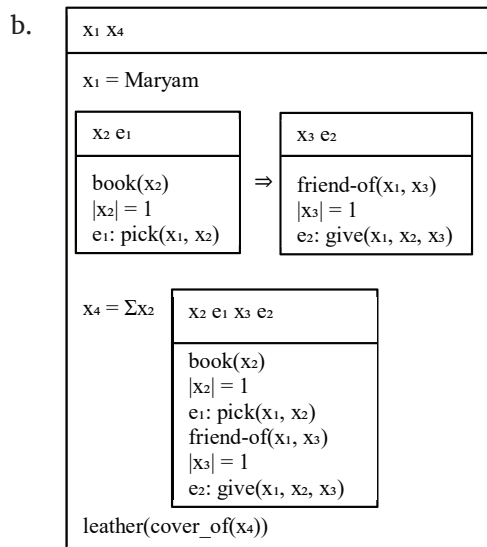
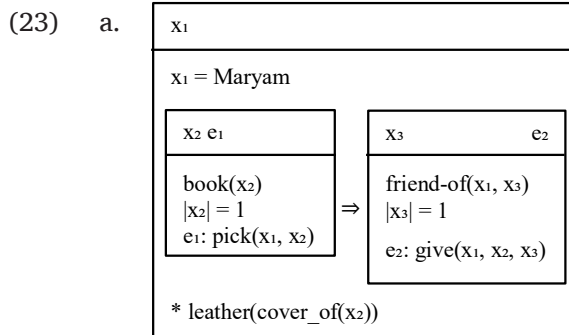
DRSs are structural representations of textual information that have truth conditions. Formally, truth conditions are expressed with respect to models of worlds consisting of a domain of entities and events that have certain properties and stand in certain relations to each other. A DRS is true with respect to such a model if there is a way to anchor all the DRs of the DRS to entities in the model such that all the conditions are satisfied in the model. For the DRS (20)(c) this means that there must be an assignment function g that maps the DRs $x_1, x_2, e_1, x_3, e_2, e_3$ to entities in the model such that the following holds:

- (21) $g(x_1)$ is the person Maryam
 $g(x_2)$ is a book
 $g(e_1)$ is an event in which $g(x_1)$ picks up $g(x_2)$
 $g(x_3)$ is a friend of $g(x_1)$
 $g(e_2)$ is an event in which $g(x_1)$ gives $g(x_2)$ to $g(x_3)$
 $g(e_3)$ is an event in which $g(x_3)$ reads $g(x_2)$,

If there is such a function g , the DRS after (19) is true in this model, otherwise false. There can be several such functions, as there might be different books or different friends of Maryam that satisfy the conditions, or different events involving the same participants (we make the somewhat artificial assumption that names are unique).

DRT provides a representation format for *donkey sentences* in which the DR of an indefinite or eventive expression is introduced in the restrictor of a conditional or a quantifier and taken up in its nuclear scope. An understanding of the anaphoric options in donkey sentences is critical for the proposal to be developed here. Consider the following example (22) and its representation (23)(a).

- (22) a. *Har vaght Maryam yek ketāb mi-kharid, oon-o be yek doost-i mi-dad.*
 every time Maryam one book DUR-buy.3SG that-OM to one friend-IDF DUR-gave3SG
 ‘Whenever Maryam bought a book, she gave it to a friend.’
- b. *#Jeld-esh charmi bood.*
 cover-3SG leather was.3SG
 ‘Its cover was of leather.’



Quantification introduces a complex DRS condition with the connector \Rightarrow . The DR for the book, x_2 , is introduced in the antecedent DRS and can be taken up in the consequent DRS, but not in subsequent sentences. The truth conditions are as follows:

- (24) The complex condition is true for an assignment g if and only if it holds that every extension of g to an assignment g' that makes the antecedent DRS true, i.e. for which it holds that $g'(x_2)$ is one book, and $g'(e_1)$ is an event in which $g'(x_1)$ picks up $g'(x_2)$, can be extended further to an assignment g'' that makes the consequent DRS true, i.e. for which it holds that $g''(x_3)$ is one friend of $g''(x_1)$, and that $g''(e_2)$ is an event of $g''(e_2)$ giving $g''(x_2)$ to $g''(x_3)$ in the model.

As a consequence of this interpretation rule, the DR x_2 is not available outside the complex DRS condition. However, certain types of anaphoric reference are in fact possible; (22)(a) could be continued as follows:

- (22) c. *Jeld-eshoon charmi bood.*
 cover-3PL leather was.3SG
 ‘Their covers were of leather.’

Kamp & Reyle (1993: 4.1.2) propose that such cases involve a process of abstraction over the DRSs of the complex DRS condition followed by a summation over one discourse referent. Applied to our example, this leads to the DRS (23)(b). The DRS construction rule licenses the introduction of a condition $x = \Sigma y [DRS \cup DRS']$ after a condition $DRS \Rightarrow DRS'$, where x is a new DR, $[DRS \cup DRS']$ is the union of the two DRSs, and y is a DR in the discourse referents of this union, with the following interpretation:

- (25) If g is an assignment, then Σx DRS is anchored to the sum of all entities d such that g can be extended to a g' , with $g'(x) = d$, such that DRS is true under g' .

This ensures that the DR x_4 is anchored to the sum of all entities d such that d is a book, and there is an event in which Maryam picks up d , and there is a friend of Maryam and an event such that Maryam gives d to that friend. Depending on the model, x_4 can be anchored to one or more books; this is responsible for the number-neutral interpretation. This DR x_4 is introduced in a position that it is accessible for subsequent update, as indicated in the last line of (23)(b).

4.3 The interpretation of BN Objects

Krifka & Modarresi (2016) interpret existential closure by a monadic DRT quantifier \exists . It receives an existential interpretation and disables the regular anaphoric uptake of any indefinite or event

expression within the scope of the quantifier. For example, (1)(a) has the syntactic form (26) and is interpreted as (30)(a) below.

- (26) *Maryam*₁ \exists [_{v,p} *t*₁ *ketāb kharid*]
 Maryam book bought
 ‘Maryam bought a book / books’

A condition of the form \exists [DRS], as it occurs in (30)(a), is interpreted as follows:

- (27) g satisfies \exists [DRS] in a model if and only if there is at least one extension g' of g that satisfies DRS in the model.

For (30)(a), the \exists -condition is true for an assignment g if and only if the following holds:

- (28) There is at least one extension g' of g such that $g'(x_2)$ is a unique book involved in the event $g'(e_1)$, the cardinality of $g'(x_2)$ is 1, and $g'(e_1)$ is an event of buying of $g'(x_2)$ by $g'(x_1)$.

The conditions $x_2 = \text{book-of}(e_1)$ and $|x_2| = 1$ express that x_2 is the unique single book related to the event e_1 . The BN *ketāb* is interpreted as a functional definite (cf. Löbner 1985), and the cardinality of the DR x_2 is 1, due to the singular feature of the count noun *ketāb* (the plural form is *ketāb-hā*). However, number neutrality can be derived from the interpretation of \exists in (27), which allows for more than one extension of the assignment g .

The sentence (1)(b), *Maryam yek ketāb kharid* ‘Maryam bought one book’, also is compatible with Maryam buying more than one book. But then the use of the number word *yek* ‘one’ introduces alternatives like *Maryam do-ta ketāb kharid* ‘Maryam bought two books’ and triggers the scalar implicature that stronger alternatives could not be truthfully uttered. The DRS quantifier \exists does not come with alternatives, hence (26) does not have this implicature.

This said, cases in which only one book was bought are semantically simpler than others, inviting an interpretation in which x_2 and e_1 can only be anchored to a single entity and event. This tendency will be counteracted by general expectations, as in sentences like *Maryam havij kharid*, ‘Maryam bought a carrot/carrots’, as it is implausible, given our world knowledge, that Maryam bought only one carrot.

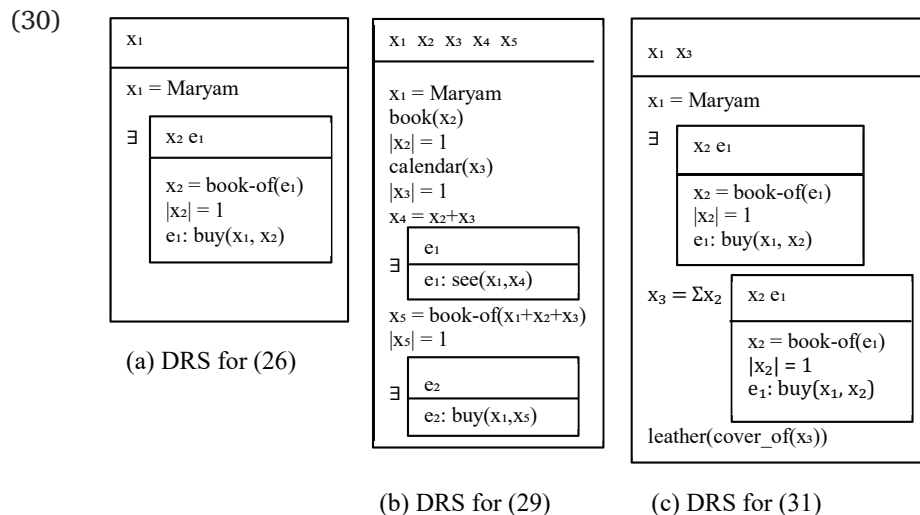
The interpretation of the BN object similar to a functional definite is made plausible by the existence of weak definites in languages like English, cf. (8), which actually mark such nominals as definites. Is also motivated by language-internal evidence, as a BN that is interpreted outside existential closure typically receives a definite-like interpretation, cf. (1c). In this case, the preceding discourse, situation or background knowledge must contain an entity with respect

to which there is a unique entity of the required kind. For example, if a book was introduced before, then *ketāb*, interpreted outside the scope of existential closure, may refer to that book, as illustrated in (29) and the corresponding DRS in (30b).

Notice that the DR for the object in the second clause is interpreted as the unique book in the antecedent DR x_4 , which is the sum of the previously introduced book x_2 and the record x_3 . Hence, a uniform interpretation of bare nouns is possible.

- (29) a. *Maryam too foroushgah yek ketāb va yek taghvim did.*
 Maryam in store one book and one calendar saw.3SG
 ‘Maryam saw a book and a calendar in the store.’
- b. *Ketāb₁-rā ∃ [_{VP} t₁ kharid]*
 book-OM bought
 ‘Maryam bought the book.’ cf. DRS in (30b)

Being outside of the scope of \exists , *ketāb-rā* cannot depend on e_1 . But it can be interpreted as relative to the sum of all salient DRs $x_1 + x_2 + x_3$, introducing the unique book among them, x_2 , and identifying it with the new DR, x_5 . In general, RA-marked bare nouns are interpreted like functional definites that identify the unique entity in their reference situation or in the world knowledge assumed to be shared by the participants of conversation.⁵ When applied to a specific DR, the possessive form, e.g. *jeld-esh/-eshoon* ‘its / their cover’, cf. (22b,c) is used.



$x_1 x_3$

$x_1 = \text{Maryam}$

\exists $x_2 e_1$

$x_2 = \text{book-of}(e_1)$
 $|x_2| = 1$
 $e_1: \text{buy}(x_1, x_2)$

$x_3 = \Sigma x_2$

$x_2 e_1$

$x_2 = \text{book-of}(e_1)$
 $|x_2| = 1$
 $e_1: \text{buy}(x_1, x_2)$

$\text{leather}(\text{cover_of}(x_3))$

⁵ This explains the use of RA-marked nouns in generic (characterizing) sentences (cf. Krifka 2001), as in *kawboy tan-bako-rā mi-javand* ‘A cowboy CHEWS tobacco.’, ‘What a cowboy does with tobacco is, they chew it.’, a quantification over situations that contain cowboys and tobacco.

Consider now the option for anaphoric uptake after (26), as in (31):

- (31) *Maryam*₁ ∃ [_{vp} *t*₁ *ketāb kharid*]. *Jeld-esh* / *-eshoon charmi bood*.
 Maryam book bought cover-3SG/3PL leather was.3SG
 ‘Maryam bought a book / books.’ ‘Its / their cover was/were of leather.’

The DRs introduced in the scope of existential closure like x_2 in (30a) are not immediately accessible. But the abstraction and summation rule (25) applies, cf. (30c), where x_3 is the sum of all books that were bought by Maryam.⁶

One consequence is that BNs are interpreted as maximal. In (30), the DR x_3 is anchored to the union of all books that Mary bought, within the described discourse universe. This might appear surprising, but maximality effects of pseudo-incorporated nominals and weak definites have been noticed in Dayal (2011) and Schwarz (2014). Maximality can be detected in the contrast between indefinite object and BN in example (32).

- (32) *Ali #(ye) khaneh dareh. Khane-ye-digari ham dareh ke ejareh mideh*.
 Ali one house has house-of-other also has that rent gives.
 ‘Ali owns a house. He also owns another house that he rents.’

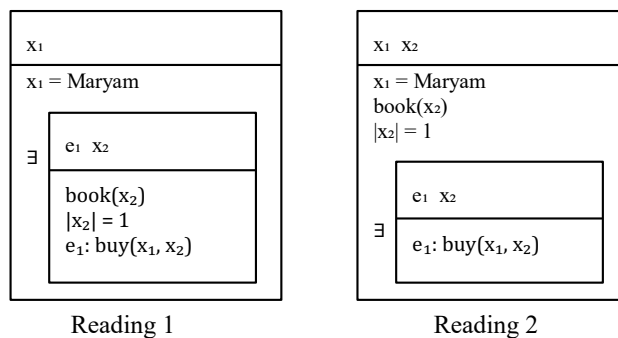
The indefinite antecedent *ye khaneh* (with unstressed, non-numeral reading of *yek*) is not interpreted as implying that Ali has only one house, whereas the BN object implies reference to the sum of all houses that Ali has.

4.4 The interpretation of YK objects and of complex predicates

YK objects are regular indefinites that are not dependent on another DR. As Fodor & Sag (1982) showed, indefinites may scope outside of island. This means that the DR of an indefinite can be introduced in the highest DRS (cf. Kamp & Reyle 1993: 288ff), leading to the following two interpretations of examples like (33):

- (33) *Maryam*₁ ∃ [_{vp} *t*₁ *yek ketāb kharid*]
 Maryam one book bought
 ‘Maryam bought a book.’

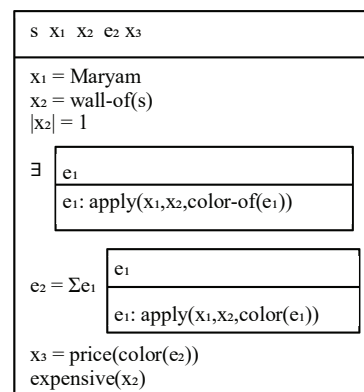
⁶ A type of summation is proposed in Schwarz (2014) in the antecedent clause, where PIN and weak definite objects lead to the formation of event kinds. In the current theory, summation only happens under anaphoric uptake.



The readings are truth-conditionally equivalent but Reading 2 introduces an accessible DR, x_2 . Notice that Reading 1 could be expressed by the simpler clause with a BN object, (30a). Hence, we should expect that Reading 1 tends to be blocked by the BN clause, following considerations of economy similar to (Fox 1998): Existential closure \exists just expresses the existence of an extension of the variable assignment that verifies the embedded DRS, not the existence of precisely one such extension. Hence the number restriction $|x_2| = 1$ expressed by *yek* is uninformative. Hence even in the absence of RA marking, indefinite singular objects are rather interpreted following Reading 2, resulting in an accessible DR x_2 . This prediction will be put to test in Section 5.

BN objects can also be interpreted as parts of complex predicates, which have an idiomatic interpretation. Regular BN objects and BN objects in complex predicates are syntactically distinct (cf. Modarresi 2014; Megerdooian 2012). One difference is that BNs in complex predicates do not allow for anaphoric uptake:

- (34) *Maryam divar-rā rang zad.*
 Maryam wall-OM color hit.3SG
 ‘Mary painted the wall.’
- #Gheimat-e-sh geroon bood*
 price-of-it expensive was.3SG
 ‘Its price was high.’
- Gheimat-e rang geroon bood*
 price-of color expensive was
 ‘The price of the paint was high.’



The expression *rang zad* ‘paint hit’ is a complex predicate with the meaning ‘paint’. The BN *rang* ‘color, paint’ cannot be taken up directly, different from BN objects considered so far. What is possible, however, is to refer to the paint by the referring expression description *gheimat-e rang* ‘the price of the paint’. This is typical for associative or “bridging” definites (cf. Charolles 1999),

which is licensed here as painting events are associated with paint. This is captured in the DRT representation above as follows: First, a DR for Mary, x_1 , and for the wall in the given situation s , x_2 , is introduced. The BN of a complex predicates refers to the object related to the event but does not introduce a DR. The complex predicate introduces an event variable, subject to existential closure as usual. The subsequent sentence can pick up this event via abstraction and summation, and the associative definite can introduce a DR that is uniquely related to that event.⁷

4.5 Predictions concerning anaphoric accessibility

Our modelling of BN objects and YK objects yields predictions about their anaphoric potential, under the assumption that directly introduced DRs with YK objects are more salient than DRs with BN objects that are introduced by abstraction and summation. Unfortunately, there is little empirical work that deals with summation anaphora (except for so-called complement anaphora, cf. Nouwen 2020 for overview). However, it is plausible that the more complex construction of DRs should reduce the availability of such antecedents. Hence hypotheses BN-0 and BN = YK, cf. (14), should be rejected, supporting BN < YK in (35), our main hypothesis. As BN antecedents do not denote implicit participants that would be only accessible by associative anaphora, hypothesis BN-DD is rejected as well.

The predictions concerning hypotheses BN-NL! and BN-NL, that BN antecedents require or prefer null (NL) anaphora, is a more complex issue. Kamp & Reyle (1993) distinguish between atomic and non-atomic DRs, where atomic DRs can be picked up by SG anaphora, and non-atomic DRs by PL anaphora. Under the assumption that YK antecedents introduce atomic DRs, they should be incompatible with PL anaphora, and should prefer SG over NL anaphora as the presupposition of SG anaphora are satisfied, following the maximize presupposition principle, cf. hypothesis YK-SG.

As for BN antecedents that involve abstraction and summation, their DRs could be either ambiguously singular or plural, or else vague, assuming number-neutral DRs (a notion also proposed in Kamp & Reyle 1993). The type of DR depends on whether the antecedent clause invokes a single object or multiple objects ('buy boat' vs. 'buy button'), or whether it lacks such biases ('buy book'). We tried to construct examples without bias for the experiments to be reported here. This predicts that BN antecedents should be picked up by NL anaphora, leading to hypothesis BN-NL in (14). However, under the single anchor preference of Section 4.3, we should expect that BN antecedents tend to introduce atomic DRs, leading to a counteracting preference of SG anaphora. As we cannot estimate the strength of the tendencies for NL vs. SG anaphora, BN-NL/SG just states that both are allowed equally; hypothesis BN-noPL states that PL anaphora are avoided.

⁷ There are also intransparent idiomatic complex predicates such as *chune zad*, lit. 'chin hit', meaning 'haggle'; they allow for no reference to the BN component.

- (35) BN < YK: BN objects allow for uptake less easily than YK objects.
 YK-SG: YK objects do not allow for PL anaphora, prefer SG over NL anaphora.
 BN-NL/SG: BN objects in non-biased contexts allow for NL and SG anaphora.
 BN-noPL: BN objects in non-biased contexts disfavor PL anaphora.
 BN > YK-PL: BN objects allow more easily for PL antecedents than YK objects.

As suggested by an anonymous referee, hypothesis BN-noPL could be due to the singular number feature of BNs. However, then we should see no difference to the singular YK nouns. Hypothesis BN > YK-PL rejects this explanation.

5 Experimental findings on anaphoric uptake

The predicted difference in saliency of DRs introduced by YK vs. BN objects is subtle and cannot be addressed introspectively or by direct observation. What can be observed is the rating of competent speakers of anaphoric uptake, the frequency of uptake and behavioral or neurophysiological measures in the processing of expression that include anaphoric uptake, which arguably can be correlated to the salience of DRs.

Frequency data using corpora is difficult to come by because saliency may depend on a number of other factors that are hard to control.⁸ Also, Persian is a highly diglossic language, there are very few corpora of spoken Persian (Mohammadi 2019 became available after our work was completed), and BN objects seem to occur less frequently in spoken language.⁹ Hence experiments is the option of choice.

In Section 5.1 we discuss relevant previous experimental studies on anaphoric accessibility for other languages. In Sections 5.2–5.6 we then present the result of five experiments that focus on observable phenomena that are plausibly related to anaphoric potential. As saliency cannot be measured directly, it is advisable to investigate different observable phenomena, with the hope that they return comparable results. Experiment 5.2 taps into processing, using self-paced reading; this neither give a supporting nor rejecting results. Experiments 5.3 to 5.5 tests explicit

⁸ This said, it is easy to find BN objects taken up by anaphora. One example with a full DP uptake, from <https://twitter.com/febrahimzade/status/1074316272735846400?lang=en>:

(i) *az ketab-foroushi-e Tahoori ketab kharidam in faal-e hafez ra gozasht la-ye*
 from book-store -EZ Tahoori book bought.1SG this poem-EZ Hafez-OM put.3SG in.between-EZ
ketab.
 book
 ‘I bought a book/books from Tahoori bookstore, they put this poem by Hafis into the book’

⁹ Faghiri & Samvelian (2015) investigate the influence of the *preceding* context on the realization of direct objects, which is different from our current interest, which focuses on the anaphoric potential of direct objects on the *subsequent* discourse.

judgements of speakers, resulting generally in a support of the hypothesis. Experiment 5.6 investigates production in a relatively natural setting, yielding clear support of the hypothesis.

5.1 Previous studies

There are a number of experimental studies on anaphora in the context of incorporation structures. In an early study, Ward et al. (1991) investigate the anaphoric potential of morphological incorporation (e.g. *deer hunting* vs. *hunt deer*) with self-paced reading and find that participants read subsequent sentences with pronouns referring to deer faster in the second condition.

Scholten & Aguilar-Guevara (2010) research the behavior of BNs, weak definites, regular definites and indefinites in Dutch. Participants had to choose between a pronoun or a definite containing the same noun as the antecedent NP. Regular indefinites were taken up far more often by pronouns than by regular definite NPs.

Oggiani (2011) looks at BNs vs. regular indefinites in Spanish, applying a technique in which participants are asked to form small texts containing constituents like *tener casa* ‘have house’ and *tener una casa* ‘have a house’. She finds that the object is picked up about only half as often when it is realized as a BN.

Law & Syrett (2017) investigate the anaphoric uptake of BNs vs. overtly marked singular or plural nouns in object position in Mandarin (singular nouns with number word ‘one’ plus classifier, plural nouns with number word ‘three’ plus classifier). Referring to Modarresi (2015), they used stimuli where world knowledge was either not biased or biased towards single or multiple entities. The second sentence contained a singular pronoun or a plural pronoun in subject position, immediately following the object in the preceding clause. An online self-paced reading task showed that anaphoric uptake of (singular) indefinite antecedents is processed more easily than uptake of BN antecedents.

5.2 Experiment 1: Self-Paced-Reading

Inspired by Law & Syrett (2018), we measured the ease of anaphoric uptake for BN vs. YK objects by self-paced reading. This taps directly into the processing of read language and should give us reliable results about the grammatical representation.

We constructed antecedent sentences with the intention to avoid bias towards a singular or a plural interpretation of the BN object and investigated cases in which the subsequent sentence had a null (NL), singular (SG) or plural (PL) anaphor, resulting in a 2×3 design (2 antecedents, 3 anaphora). A sample item in six conditions is presented in (36).

- (36) *Leili az ketāb-foroushi ketāb / yek ketāb khærid, va ba deghat*
 Leili from book-store book / one book bought.3SG and carefully
kado-∅ /-esh /-eshoon kard, va be khaneh-ye-doost-a-sh raft.
 wrapped-∅ / it / them did and to house-EZ-friend-EZ-her went.3SG
 ‘Leili bought a book/books from bookstore, carefully wrapped (∅-it-them) and went to
 the house of her friend.’

As for the hypotheses in (14), BN-0 predicts slowing down with BN objects whereas BN = YK predicts no such slowdown (rather a boost as BN objects are shorter than YK objects). BN-NL! would be supported if BN followed by NL is processed faster than BN followed by SG/PL, in case this difference does not obtain for YK followed by NL vs. YK followed by SG/PL.

64 native Persian speakers participated in an experiment designed with Ibx-Farm, both online and in the lab. There was no significant difference between the two groups; we report the overall results for both groups. The stimuli consisted of 48 sentence items similar to (36). We tried to construct these sentences with frequent words that are encountered in every-day Persian conversation.¹⁰ Comprehension questions after more than half of the trials checked for attention of the participants. One example of a comprehension question following (36) is *Ki ketāb kharid?* ‘Who bought a book/books?’ followed by a multiple-choice task. In the items, the anaphoric element does not follow the antecedent immediately and is not realized in the beginning of the sentence. Furthermore, antecedent and item were always in object position. There were 48 trials per person. Trial types were presented in a Latin-square design with 6 conditions of similar structure in six different lists. Each participant saw each sentence under just one condition.

Participants read the sentences fragment by fragment in Persian orthography, from right to left, by pressing the space key. Response times between key presses were recorded. Participants answered a comprehension question, with a high accuracy rate (mean = 94.23%), indicating that they paid attention.

We had to delete data from six participants because their mean reading times were suspiciously slow or high. As usual, the measured reading times had a thick right tail, so we used a boxcox transformation to get a normally distributed response variable. A log likelihood profile (R-package) showed the best value for the boxcox parameter to be $\lambda = 0.4$. The results are

¹⁰ One reviewer asked for an estimation of the frequency of BN objects vs. other objects, in particular YK-marked objects, which might affect anaphoric uptake if BN objects are a rare phenomenon. This is not the case; a count of objects nouns in literary narrative texts revealed that while RA-marked objects occur most frequently, YK-marked objects and BN objects occur about equally often. BN objects occur particularly often as complements of prepositions.

shown in **Figure 1** from the word before the anaphoric expression, at the word containing the anaphoric expression, and the following two words.

We saw an effect of word length of the anaphora (NL < SG < PL) across BN/YK antecedents. The difference between BN followed by SG vs. YK followed by SG was not significant. We concluded that self-paced reading is not a suitable task for investigating the processing of anaphoric reference, at least not in the setup we have chosen.

5.3 Experiment 2: Acceptability Judgements

In a second experiment, we used the same setup and materials as in Experiment 1, except that the reading tasks were followed by an acceptability question on a Likert Scale from 1 (poor) to 7 (good). We wanted that the self-paced readers pay more attention to the content of the items and have additional acceptability data. There were 38 participants, different from the ones in Experiment 1, partly in a lab and partly on-line. We got comparable results as in

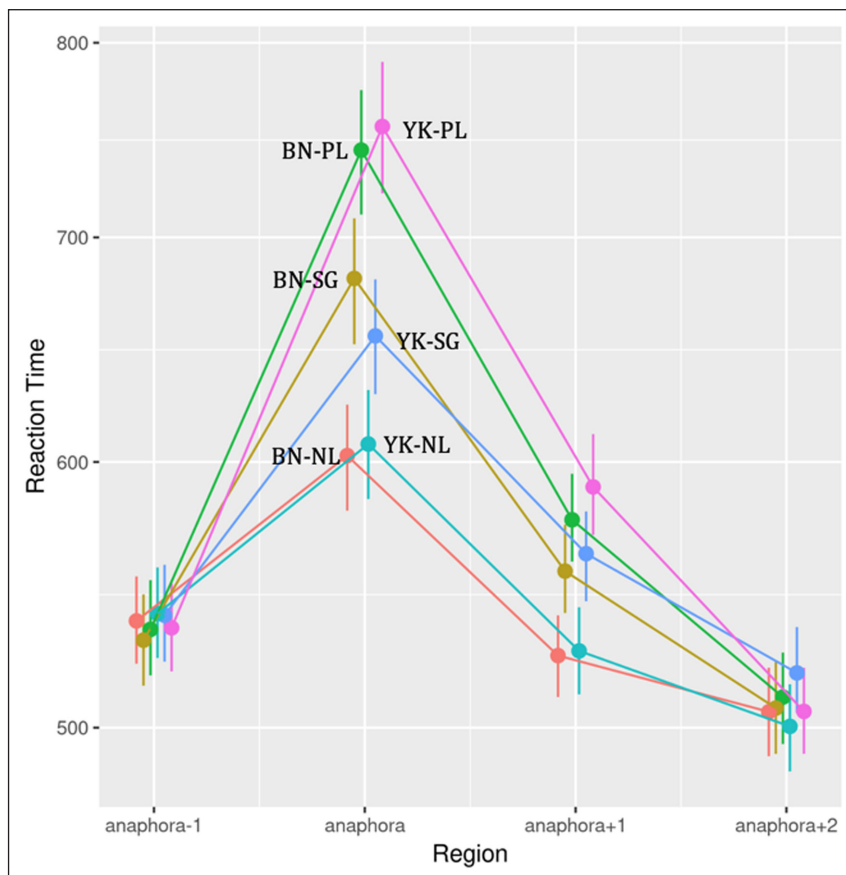


Figure 1: Results of self-paced reading test.

Experiment 1 for the reading times and do not report them here. As the ratings of participants differed widely (some judging sentences generally better, others generally worse), the judgements of each participant were z-transformed. **Figure 2** gives a whisker diagram of the z-transformed data (e.g., +1 indicates one standard deviation difference from the mean); **Figure 3** then indicates the means of the indicated continuations. The cases are presented in the same order, where NL/SG combines the NL and SG cases. Cases with BN antecedents are given in white, cases with YK antecedents are given in grey; the order of presentation is the same in the two figures.

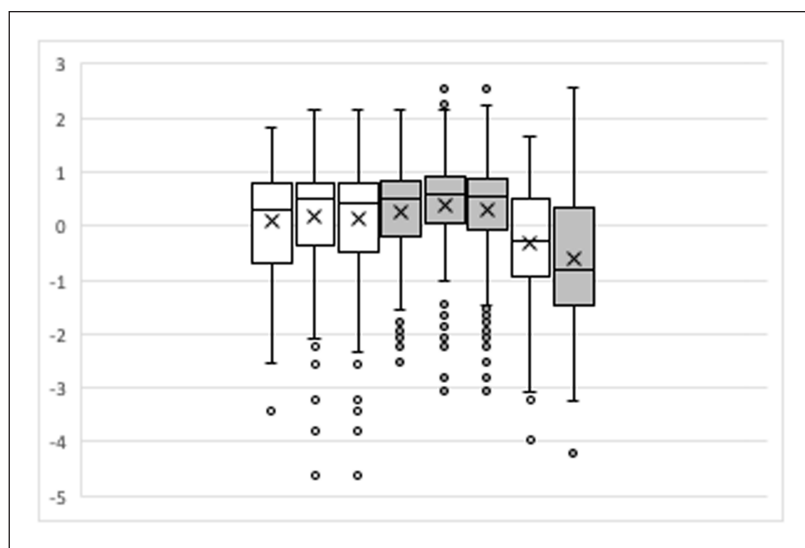


Figure 2: z-transformed ratings, with means indicated by \times ; y-axis: standard deviations from mean; YK antecedents in grey.

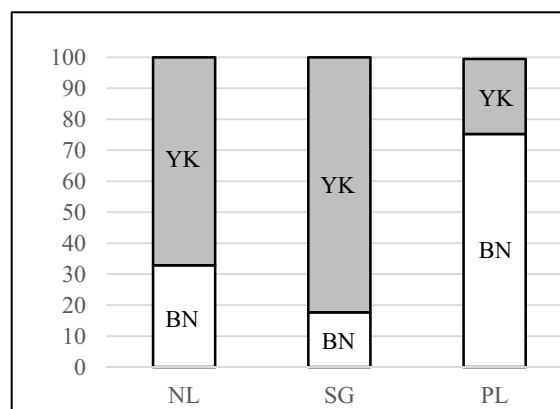


Figure 3: Means of z-transformed ratings, relevant significant differences indicated; y-axis: standard deviations from mean; YK antecedents in grey.

Starting with the non-PL cases, we find that anaphora to BN antecedents were generally rated worse than anaphora to YK antecedents (BN-NL/SG vs. YK-NL/SG, non-paired t-Test assuming equal variances: $p < 0.005$), contra hypothesis BN = YK. When comparing the non-transformed averages of each participant across items using a paired t-test, the difference is also significant ($p < 0.05$). The means of BN-NL/SG was positive, contra hypothesis BN=0. Together, this result supports the main hypothesis, BN < YK. BN-NL vs. BN-SG is not significant, consistent with hypothesis BN-NL/SG; YK-NL vs. YK-SG is approaching significance ($p = 0.064$), supporting hypothesis YK-SG. As for the PL cases, notice that they both are judged negatively (significant differences to all non-PL cases, not indicated in graph), supporting hypothesis YK-SG and BN-noPL. In addition, YK-PL is judged more negatively than BN-PL ($p < 0.001$), supporting hypothesis BN > YK-PL, even though we found it surprising how negatively BN-PL cases were actually judged.

5.4 Experiment 3: Anaphora Choice

In the third experiment we tested the anaphoric potential of BN and YK objects in a forced-choice selection of anaphoric expressions, a controlled production experiment. This should reflect linguistic competence more directly than a rating experiment. We wanted to test whether there is a requirement or preference of BN antecedents to select for NL anaphora, as claimed in hypotheses BN-NL! and BN-NL, or no such preference, after hypothesis BN-NL/SG.

Participants were presented with a sentence containing a BN or YK object, and a continuation sentence followed by a blank to be filled with multiple choices of NL, SG, and PL anaphora. As a sample item, consider (37).

- (37) *Banna too sakhteman divar / yek-divar sakht, modati baad ...*
 builder in building wall / one wall built, sometime later ...
 The builder constructed a wall in the building, but sometimes later...'

kharab-Ø kard *kharab-esh kard* *kharab-eshoon kard*
 destroy-Ø did.3SG destroyed-it did.3SG destroyed-them did.3SG

We constructed 36 test items with two conditions and no obvious bias to singular or plural interpretation, including 8 fillers. There were 153 native Persian speakers that voluntarily participated in this experiment using an online survey platform, answering an average of 16 questions. The stimuli were presented in six different lists (participants also participated in Experiment 4, but it was made sure that no participant saw the same sentence twice). Each list included both conditions, with an average of four fillers; the items were randomized using Latin square design. The results are indicated in **Figure 4**.

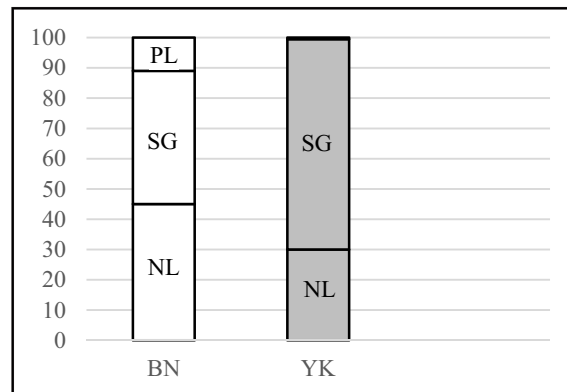


Figure 4: % Anaphora choice (NL/SG/PL) dependent on given antecedent (BK/YK).

BN antecedents are taken up by NL or SG anaphora about equally often, but significantly less often by PL anaphora, supporting hypothesis BN-NL/SG contra hypotheses BN-NL! and BN-NL (Chisquare test, $p \approx 0$). The infrequent uptake by PL anaphora supports hypothesis BN-noPL. In comparison, YK nouns are significantly more frequently taken up by SG than by NL anaphora, supporting hypothesis YK-SG. PL anaphora are taken up quite rarely. If they are taken up, then only with BN antecedents, supporting hypothesis BN > YK-PL.

5.5 Experiment 4: Antecedent Choice

Experiment 3 did not show whether the participants favored the use of BNs as antecedents of anaphora because the antecedents were given. We reversed the design and investigated the forced choice of antecedents (BN vs. YK object), when the anaphor in the subsequent sentence is given (as NL, SG or PL). Otherwise, the stimuli were the same as in Experiment 3.

- (38) *Banna too sakhteman* □ *divar* □ *yek-divar sakht, modati baad ...*
 builder in building wall one wall built, sometime later ...
- kharab-Ø kard* / *kharab-esh kard* / *kharab-eshoon kard*
 destroy-Ø did.3SG destroyed-it did.3SG destroyed-them did.3SG

There were 30 items with three conditions, including 6 fillers. The same 153 native Persian-speakers as in Experiment 3 participated in this experiment as the second part of the experiment. The stimuli were presented in six different lists. Each list included all the three conditions in randomized order, including an average of four fillers in each list, using a Latin square design. After reading the whole sentence, the participants had to choose whether the BN or the YK noun was the most appropriate antecedent. Results are presented in **Figure 5**.

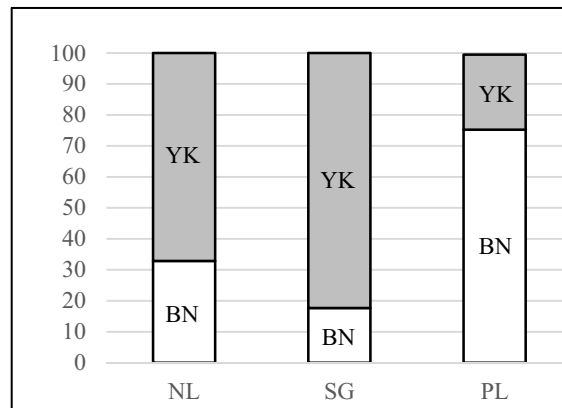


Figure 5: % Antecedent choice (BN/YK) dependent on given anaphora (NL/SG/PL).

This experiment shows that YK objects make better antecedents, except in the case of a PL anaphor, which leads to a semantic clash. We were surprised that YK objects were selected at all. The reason could be that YK objects in general make better antecedents, but we suspect that the task of going back in the text, choosing an antecedent, reading the text, choosing another antecedent, reading the text in this version, and selecting the better option was quite complex, at least for some participants.¹¹ Focusing on the non-PL cases, the experiment clearly rejects BN = YK, that BN and YK nouns are equal in their anaphoric potential (Chisquare, $p < 0.001$). It definitely rejects hypothesis BN-0, the widespread assumption that BN cannot serve as antecedents.¹² thus supporting the main hypothesis BN < YK.

5.6 Experiment 5: Free Sentence Completion

In a final experiment, we investigated which anaphoric forms are used by participants spontaneously in a context favoring anaphoric uptake, contrasting BN and YK-marked antecedents. This task does not investigate the reflection of participants about language, but asked that the participants perform a natural linguistic activity. In particular, it also leaves open the option of no anaphoric uptake at all. A sample item is (39).

- (39) *Negar dar ketābkhooneh rooznameh / yek rooznameh khoond va _____*
 Negar in library newspaper / one newspaper read.3SG and
 ‘Negar read newspaper / a newspaper in the library and _____’

¹¹ On a suggestion of a reviewer, we took YK-PL reactions as an exclusion criterion and removed all 11 of 51 participants that gave two or more such answers. This did not change the overall results for NL and SG anaphora, which were: BN-NL 33% YK-NL 67%, BN-SG 16%, YK-SG 84%, BN-PL 86% and YK-PL still remaining 14%.

¹² Even if we assume an error rate of participants of 20%, BN-0 is strongly refuted ($p < 0.00001$).

(39) does not have a specific bias towards a singular or plural interpretation, as it is as likely to read one or more than one newspaper in a library. The stimuli consisted of 24 items with two conditions, randomized in a Latin Square design in different lists. There were altogether 252 participants that took part in an online experiment. Participants read an average of three sentences (in the neutral context; there were also other items for singular and plural contexts, as well as fillers) and were asked to type a suitable continuing sentence. We collected 754 test items after exclusion of incomplete answers. Every sentence was analyzed separately to see if and how the participants referred back to the antecedent object noun. Naturally, there was a greater variety in the anaphoric responses. The results in **Figure 6** visualize NL anaphora, singular anaphoric reference with pronouns or clitics (Pro-SING), singular anaphoric reference with full DPs (Full DP-SING), plural anaphoric reference with pronouns or clitics (Pro-PLUR) and plural anaphoric reference with full DPs (Full DP-PLUR). Associative plurals and reference to kinds were very rare and are not reported here.¹³

We see that BN objects are taken up by anaphora more frequently than not (BN objects are not taken up only about 36% of the time), clearly disproving hypothesis BN-0. But YK objects are taken up more frequently than BN objects, arguing against hypothesis BN = YK.

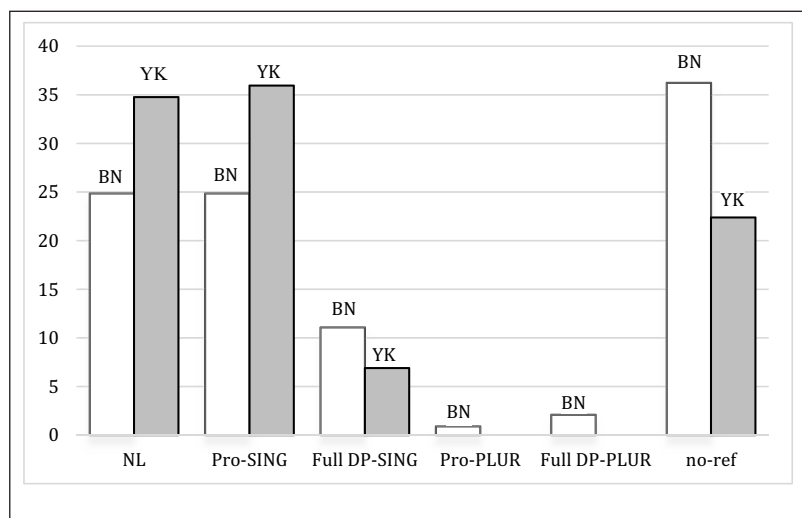


Figure 6: Free completion task; x-axis: anaphoric device.

¹³ To be sure, associative anaphora do occur in Persian, as in other languages. For the experimental task of sentence continuation, participants did not employ this device because it requires the introduction of new DRs that are licensed by world knowledge, which requires additional effort. We would also like to remark that an associative anaphora analysis of anaphoric uptake by pronouns is implausible. Persian has no grammatical gender, hence pronouns are semantically impoverished compared to languages like German and even English. Associative anaphora of the type *Leili got married. He is nice.* are impossible in Persian.

However, this difference is not quite significant (Chisquare: $p < 0.07$). Together, this is strong support of the main hypothesis, $BN < YK$. Anaphoric uptake of BN objects is rarely by full DPs, against the associative anaphora hypothesis, $BN-DD$. BN antecedents are about equally often picked up by NL and SG anaphora, disproving hypotheses $BN-NL!$ and $BN-NL$ and supporting hypothesis $BN-NL/SG$. Interestingly, YK antecedents are about equally taken up by NL and ProSING antecedents, not supporting the second part of $YK-SG$, in contrast to experiment 3. PLUR anaphora occur very rarely, supporting hypothesis $BN-noPL$; the few cases are anaphoric to BN antecedents, supporting hypothesis $BN > YK-PL$.

6 Discussion

The experimental findings clearly disprove hypothesis $BN-0$ and support hypothesis $BN < YK$: BN objects make surprisingly good antecedents, though not quite as good as YK objects. This cannot be due to associative anaphora, which occur rarely (hypothesis $BN-DD$ disproved), and it cannot be due to a process that specifically favors null anaphora (hypotheses $BN-NL!$, $BN-NL$ disproved).

As we have seen, linguists that work on their native language often denied that BN (or PIN) objects allow for anaphoric uptake. Why is there this difference with our experimental results? We suspect that this is due to a problem with the introspective access to linguistic data: When researchers ponder over the anaphoric possibilities of a PIN object vs. a regular object, the anaphoric potential of the latter is more obvious, leading to the impression that the anaphoric potential of the former is nearly absent. This stresses the importance of empirical research that takes into account different effects of saliency and the intuitions and productions of a larger number of speakers. It also casts doubt on the notion that grammaticality distinctions are binary, and supports approaches to grammar that accept gradient acceptability judgments.

Not every experimental procedure gave interpretable results. In particular, the self-paced reading experiment in Section 5.2 was unsuccessful in the sense that the results depended on an orthogonal factor, the length of the anaphoric expression. The other methods – ratings, forced production of anaphor and antecedent, and free sentence completion – showed results comparable with each other. These procedures can be seen to be different measures of anaphoric potential, a rather abstract notion, and hence in combination strengthen our inference about the anaphoric potential.

In particular, the rating experiment showed that BN antecedents, while slightly worse than YK antecedents, were judged rather good (**Figure 2**). The antecedent choice experiment showed that BN antecedents were selected quite often, even though less often than YK antecedents (**Figure 4**). And the free completion task, in which participants could have avoided of anaphoric reference, showed that in most of the cases, they *did* pick up antecedent BNs (**Figure 6**). The anaphora

choice experiment (**Figure 3**) showed that BN antecedents were picked up about as often by overt SG pronouns as with null pronouns. This shows that there is no special requirement that BN (or PIN) antecedents are picked up by null anaphora, as claimed by Farkas & de Swart (2003) and, to some extent, by Modarresi (2015).

Our findings support the analysis of Krifka & Modarresi (2016) in contrast to other theories. It predicts anaphoric uptake of BN antecedents to be more complex, and it offers an explanation for this as it involves summation over the existential closure. It should be stressed that the operation of summation was not invented for the current purpose but is part of the standard repertoire of DRT (cf. Kamp & Reyle 1993).

Krifka & Modarresi (2016) predicts that the anaphoric potential of BNs should be similar to other cases that involve abstraction and summation, as for example in donkey sentences, cf. (31). Unfortunately, there is no experimental or corpus data available on such uptakes. It would be a natural next step to investigate the ease of such anaphoric relations. Donkey sentences involve the cumulation of two sub-DRSs, the antecedent DRS and the consequent DRS, which appears to be a more complex operation than just referring to one sub-DRS, as in the case of BNs. Consequently, we expect that anaphoric reference to BNs is slightly easier than anaphoric reference to antecedents within a donkey sentence.

There are a number of follow-up questions arising from our findings that we did not address in the current article: (1) The anaphoric potential of antecedent clauses with a bias towards a singular or plural interpretation; we expect that in the latter case, plural anaphora are used more frequently. (2) The anaphoric potential of different semantic classes of verbal predicates; we expect that this affects BN and YK antecedents similarly. (3) The anaphoric potential of BN objects vs. transparent complex predicates; we expect that the latter is lower, and favors associative anaphora by definite descriptions. (4) The experimental testing of the maximality effect in (32). (5) The effect of plural marking on the object and (6) the effect of the dependent indefinite marker *-i* that occur with objects with and without RA marking (cf. Modarresi 2014 for discussion). Lastly, looking beyond Persian, it remains to be seen whether syntactic phenomena that have been identified as PIN in other languages have the same anaphoric potential, and hence are open to a similar analysis as presented here for Persian bare noun objects.

Acknowledgements

We thank the audiences at presentations at the workshop “Sorting out the concepts behind definiteness” at the DGfS Annual Meeting 2019 in Bremen, the 42nd conference Generative Linguistics in the Old World (GLOW 42) in Oslo and the Second North American Conference in Iranian Linguistics (NACIL2) in Tucson, all in 2019. We thank our collaborators, in alphabetical order, Nina Hosseini, Hakimeh Rezayie, and Fahimeh Taheri, and Werner Frey for important comments. And in particular, we thank the three anonymous reviewers that provided many detailed hints concerning the content and presentation of this article. Of course, we cannot be sure that they agree with the final result.

Funding Information

Work on this paper was funded by DFG Grant KR951/10-1, “Anaphoric Potential of Incorporated Nominals and Weak Definites” (ANAPIN).

Competing Interests

The authors have no competing interests to declare.

References

- Aguilar-Geuvara, Ana & Zwarts, Joost. 2010. Weak definites and reference to kinds. *SALT* 20. 179–196. DOI: <https://doi.org/10.3765/salt.v20i0.2583>
- Asudeh, Ash & Mikkelsen, Line Hove. 2000. Incorporation in Danish: implications for interfaces. In Cann, Ronnie, et al. (eds.), *Grammatical interfaces in HPSG*, 1–15. Stanford, CA: CSLI Publications.
- Barjasteh, Darab. 1983. Morphology, syntax and semantics of Persian compound verbs: A lexical approach. *PhD diss, University of Illinois*.
- Borik, Olga & Gehrke, Berit. 2015. An introduction of the syntax and semantics of pseudo-incorporation. In: Borik, Olga & Gehrke, Berit (eds.), *The syntax and semantics of pseudo-incorporation*, 1–46. Leiden: Brill. DOI: <https://doi.org/10.1163/9789004291089>
- Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.
- Browning, Maggie A., & Karimi, Ezat. 1994. Scrambling to object positions in Persian. In: Corver, Norbert & van Riemsdijk, Henk (eds.), *Studies on scrambling*, 61–100. Berlin: Walter de Gruyter.
- Carlson, Greg & Sussman, Rachel. 2005. Seemingly indefinite definites. In Kepser, Stephan & Reis, Marga (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 71–85. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110197549.71>
- Charolles, M. 1999. Associative anaphora and its interpretation. *Journal of Pragmatics* 31. 311–326. DOI: [https://doi.org/10.1016/S0378-2166\(98\)00070-8](https://doi.org/10.1016/S0378-2166(98)00070-8)

- Chung, Sandra & Ladusaw, William. 2004. *Restriction and saturation*. Cambridge, MA: MIT Press. DOI: <https://doi.org/10.7551/mitpress/5927.001.0001>
- Cohen, Ariel & Erteschik-Shir, Nomi. 2002. Topic, focus, and the interpretation of bare plurals. *Natural Language Semantics* 10. 125–165. DOI: <https://doi.org/10.1023/A:1016576614139>
- Dayal, Veneeta. 2003. Bare nominals: Non-specific and contrastive readings under scrambling. In: Karimi, Simin (ed.), *Word order and scrambling*, 67–90. Oxford: Blackwell Publishing. DOI: <https://doi.org/10.1002/9780470758403.ch4>
- Dayal, Veneeta. 2011. Hindi pseudo-incorporation. *Natural Language and Linguistic Theory* 29. 123–167. DOI: <https://doi.org/10.1007/s11049-011-9118-4>
- Dayal, Veneeta & Sağ, Yağmur. 2019. Determiners and Bare Nouns. *Annual Review of Linguistics* 6. 173–194. DOI: <https://doi.org/10.1146/annurev-linguistics-011718-011958>
- Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.
- Espinal, M. Teresa & Cyrino, Sonia 2017. On weak definites and their contribution to event kinds. In Fernández-Soriano, Olga, et al. (eds.), *Boundaries, Phases, and Interfaces. Case studies in honor of Violeta Demonte*, 129–149. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.239.07esp>
- Espinal, M. Teresa & McNally, Louise. 2011. Bare nominals and incorporating verbs in Spanish and Catalan. *Journal of Linguistics* 47. 87–128. DOI: <https://doi.org/10.1017/S0022226710000228>
- Faghiri, Pegah & Samvelian, Pollet. 2015. How much is determined by syntax? An empirical approach to the position of the direct object in Persian. *9th Iranian Conference on Linguistics, Allameh Tabataba'i University*, 1419–1435.
- Farkas, Donka & de Swart, Henriëtte. 2003. *The semantics of incorporation: From argument structure to discourse transparency*. Stanford: CSLI Publications.
- Fox, Danny. 1998. *Economy and semantic interpretation)— A study of scope and variable binding*. Doctoral dissertation. Cambridge, MA: MIT.
- Frey, Werner. 2010. \bar{A} -movement and conventional implicatures: About the grammatical encoding of emphasis in German. *Lingua* 120. 1416–1435. DOI: <https://doi.org/10.1016/j.lingua.2008.09.016>
- Ganjavi, Shadi. 2007. *Direct Object in Persian*, Doctoral dissertation, University of Southern California.
- Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia Linguistica* 57. 47–74. DOI: <https://doi.org/10.1111/1467-9582.00099>
- Ghomeshi, Jila. 2008. Markedness and bare nouns in Persian. In Karimi, Simin, et al. (eds.), *Aspects of Iranian linguistics*, 85–112. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Goldberg, A. E. 1996. Words by default: Optimizing constraints and the Persian complex predicate. In *Annual Proceedings of the Berkeley Linguistic Society* 22. 132–146. Berkeley. DOI: <https://doi.org/10.3765/bls.v22i1.1360>
- Gussenhoven, Carlos. 1983. Focus, mode, and the nucleus. *Journal of Linguistics* 19. 377–417. DOI: <https://doi.org/10.1017/S0022226700007799>

- Hincha, Georg. 1961. Beiträge zu einer Morphemlehre des Neupersischen. *Der Islam* 37. 137–201. DOI: <https://doi.org/10.1515/islam.1961.37.1-3.136>
- Jacobs, Joachim. 1991. Focus ambiguities. *Journal of Semantics* 8. 1–36. DOI: <https://doi.org/10.1093/jos/8.1-2.1>
- Kahnemuyipour, Arsalan. 2009. *The Syntax of Sentential Stress*. Oxford, UK: Oxford University Press.
- Kamali, Beste. 2015. Caseless direct objects in Turkish revisited. In Meinunger, André (ed.), *Byproducts and side effects: Nebenprodukte und Nebeneffekte (ZAS Papers in Linguistics 58)*, 107–123. Berlin: ZAS. DOI: <https://doi.org/10.21248/zaspil.58.2015.430>
- Kamp, Hans & Reyle, Uwe. 1993. *From discourse to logic. Introduction to model theoretic semantics of natural language, formal logic, and Discourse Representation Theory*. Dordrecht: Kluwer.
- Karimi, Simin. 2003. On object positions, specificity and scrambling in Persian. In: Karimi, Simin (ed.), *Word order and scrambling*, 91–124. Oxford: Blackwell Publishing. DOI: <https://doi.org/10.1002/9780470758403.ch5>
- Karimi, Simin & Smith, Ryan Walter. 2020. Another look at Persian *rā*. In: Larson, Richard & Moradi, Sedigheh & Samiiian, Vida (eds.), *Advances in Iranian Linguistics*, 155–172. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/cilt.351.09kar>
- Krifka, Manfred. 2001. Non-novel indefinites in adverbial quantification. In: Condoravdi, Cleo & de Lavalette, Gerard Renardel (eds.), *Logical perspectives on language and information*, 1–40. Stanford: CSLI Press.
- Krifka, Manfred & Modarresi, Fereshteh. 2016. Number neutrality and anaphoric uptake of pseudo-incorporated nominals in Persian (and weak definites in English). *Semantics and Linguistic Theory (SALT)* 26. 874–891. DOI: <https://doi.org/10.3765/salt.v26i0.3919>
- Law, Jess H.-K., & Syrett, Kristen. 2017. Experimental evidence for the discourse potential of Mandarin. *NELS* 47.
- Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4. 279–326. DOI: <https://doi.org/10.1093/jos/4.4.279>
- Margetts, Anna. 2008. Transitivity discord in some Oceanic languages. *Oceanic Linguistics* 47. 30–44. DOI: <https://doi.org/10.1353/ol.0.0004>
- Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language and Linguistic Theory* 19. 153–197.
- Massam, Diane. 2009. Noun Incorporation: Essentials and Extensions. *Language and Linguistics Compass* 3. 1076–1096.
- McNally, L. 1995. Bare plurals in Spanish are interpreted as properties. In Grammar, Formal & Morrill, Glyn & Oehrle, Richard (eds.), *Formal Grammar*, 197–212. Barcelona: Polytechnic University of Catalonia.
- Megerdooimian, Karine. 2012. The status of the nominal in Persian complex predicates. *Natural Language and Linguistic Theory* 30. 179–216.

- Modarresi, Fereshteh. 2010. Persian Bare Singulars, The role of Information Structure. Actes du congrès annuel de l'Association canadienne de linguistique 2010. *Proceedings of the 2010 annual conference of the Canadian Linguistic Association*.
- Modarresi, Fereshteh. 2014. *Bare nouns in Persian: Interpretation, grammar, and prosody*. Doctoral dissertation. Humboldt Universität zu Berlin.
- Modarresi, Fereshteh. 2015. Discourse properties of bare noun objects. In Borik, Olga & Gehrke, Berit (eds.), *The syntax and semantics of pseudo-incorporation*, 189–221. Leiden: Brill. DOI: https://doi.org/10.1163/9789004291089_007
- Mohammadi, Ariana Negar. 2019. Corpus of Conversational Persian Transcripts LDC2019T11. Web Download. Philadelphia: Linguistic Data Consortium.
- Nemati, Fatemeh. 2010. Incorporation and complex predication in Persian. *LFG10 Conference*, 395–415. Stanford: CSLI Publications.
- Nouwen, Rick. 2020. E-type pronouns: congressmen, sheep and paychecks. In: (ed.), *The Wiley Blackwell Companion to Semantics*. DOI: <https://doi.org/10.1002/9781118788516.sem091>
- Öztürk, Balkiz. 2004. Is there agent incorporation? *CLS* 40. 279–289.
- Poesio, Massimo. 1994. Weak definites. *SALT* 4. 282–299. DOI: <https://doi.org/10.3765/salt.v4i0.2465>
- Samvelian Pollet & Pegah Faghiri. 2014. Persian complex predicates: How compositional are they? *Semantics-Syntax Interface* 1. 43–74.
- Scholten, Julien & Aguilar-Guevara, Ana. 2010. Assessing the discourse referential properties of weak definite NPs. *Linguistics in the Netherlands* 41. 115–128. DOI: <https://doi.org/10.1075/avt.27.10sch>
- Schwarz, Florian 2013. Two kinds of definites cross-linguistically. *Language and Linguistic Compass* 7(10). 534–559. DOI: <https://doi.org/10.1111/lnc3.12048>
- Schwarz, Florian 2014. How weak and how definite are weak indefinites? In: Aguilar-Guevara, et al. (eds.), *Weak Referentiality*. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.219.09sch>
- Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. In: Aguilar-Guevara, Ana & Pzas Loyo, Julia & Maldonado, Violeta Vázquez-Rojas (eds.), *Definiteness across languages*, 1–38. Berlin: Language Science Press.
- Selkirk, Elisabeth O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, Mass.: MIT Press.
- van Geenhoven, Veerle. 1992. Noun incorporation from a semantic point of view. *BLS* 18. 453–466. DOI: <https://doi.org/10.3765/bls.v18i1.1580>
- van Geenhoven, Veerle. 1998. *Semantic Incorporation and Indefinite Description*. Standford: CSLI Publications.
- von Heusinger, Klaus & Jaklin Kornfilt. 2005. The case of the direct object in Turkish: Semantics, syntax and morphology. *Turkic Languages* 9. 3–44.

Ward, Gregory et al. 1991. A pragmatic analysis of so-called anaphoric islands. *Language* 67. 439–473. DOI: <https://doi.org/10.2307/415034>

Windfuhr, Gernot L. 1979. *Persian Grammar: History and State of Its Study*. The Hague: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110800425>

Yanovich, Igor. 2008. Incorporated nominals as antecedents for anaphora, or How to save the thematic arguments theory. *University of Pennsylvania Working Papers in Linguistics* 14. 367–3.

