ELM

Pseudo-Incorporated Antecedents and Anaphora in Persian:
The Influence of Stereotypical Knowledge

Fereshteh Modarresi and Manfred Krifka[*]

Abstract. There are different theories about the nature of pseudo-incorporated nouns (PINs), which feature a non-specific, number-neutral interpretation. For a proper analysis it is crucial to take their anaphoric potential into account. This paper investigates if and how PINs introduce discourse referents, with evidence from Persian, and which theory matches this behavior best. We report on experiments in which the stereotypical enrichment of the number-neutral interpretation was systematically varied with two types of biases — towards a singular or a plural interpretation — and in the neutral case, when such a bias is lacking. The results of the experiments are compatible with Krifka & Modarresi (2016), which considers PIN objects as dependent singular definites (similar to weak definites) within existential closure over an event variable.

Keywords. semantics; pragmatics; psycholinguistics; anaphoric reference; weak definites; pseudo-incorporation.

1. Pseudo-Incorporated Nominals (PINs). Pseudo-incorporated nominals (PINs) have been identified by Massam (2001) as arguments of verbal predicates that exhibit certain syntactic and semantic properties. As for their syntax, they exhibit certain restrictions with respect to their prosodic and syntactic independence but are not fully morphologically incorporated, as in *Mary went to church* in contrast to *Mary was a regular churchgoer* and *Mary went to a church.* As for their semantics, PINs are unspecific and number-neutral, as in *the people in the town went to church* – no particular church is referred to, in fact there might be more than one church. In their interpretation, PINs correspond to weak definites as *Mary was taken to the hospital*, which are expressed by syntactically reduced forms in some languages (e.g., in German *ins Hospital* vs. *in das Hospital*, cf. Schwarz (2014)). PINs may be realized in different ways in different languages, and they may be more or less prominent in certain languages (see Borik & Gehrke 2015, Massam 2017 and Chung & Ladusaw 2020 for recent treatments).

The current paper deals with direct object arguments in Persian, which has a direct object marker in form of a postposition *-rā* that triggers a specific or definite interpretation with bare nominals, cf. (1). Indefinite DPs, as marked with the indefinite number word or article *yek*, can occur with or without *rā*, cf. (1). Bare nominals without *-rā* are illustrated in (1).

(1) a. Sara ketāb rā   kharid.   b. Sara yek ketāb rā   kharid.
        Sara book OM bought   Sara one book OM bought
        'Sara bought the book   'Sara bought a book'

---

c. Sara yek ketāb kharid.
    Sara one   book bought
    'Sara bought a book'

d. Sara   ketāb   kharid.
    Sara one book bought
    'Sara bought an book / books.'

Bare objects without *-rā* (d) have a number-neutral interpretation; they are also interpreted as non-specific. As Modarresi (2014, 2015) shows, they also have the other defining properties of PINs: they form one prosodic domain with the verb, they cannot undergo scrambling but only focus movement, and can be expanded by modifiers. Hence they can be analyzed as instances of PINs.

There are a number of different theories about the nature of PINs. As for their semantics, they have been analyzed as referring to kinds (cf. Ghomeshi 2008, Aguilar-Guevara & Zwarts 2010, Dayal 2011), denoting properties (cf. McNally & van Geenhoven 2005, Dobrovie-Sorin & Giurgea 2015), and involving a restrictive function-argument composition (cf. Ladusaw & Chung 2003, Espinal & McNally 2011). Other accounts have been proposed as well.

For a proper analysis of the semantics of PINs it is crucial to take their anaphoric potential into account, i.e. their ability to introduce discourse referents (DRs) that can be picked up by anaphoric expressions in subsequent discourse. The three theoretical options mentioned above predict that PINs are not accessible for anaphoric uptake without further mechanisms, as kinds, properties, and semantic restriction would not result in object-level DRs. They predict that PINs are "discourse opaque". There are other theories of the interpretation of PINs that predict that anaphoric uptake is possible. Modarresi (2014) assumed that PINs in Persian introduce number-neutral DRs, thus capturing their number-neutral interpretation and an assumed preference for uptake by null anaphora, which are number-neutral as well. According to her, PINs are "discourse transparent" (cf. also van Geenhoven, 1998 for incorporation in Greenlandic Eskimo). Krifka & Modarresi (2016) proposed that PINs introduce DRs with restricted scope that can be extended by an operation of abstraction and summation, allowing for a restricted anaphoric uptake. In a term introduced by Farkas & de Swart 2003, PINs are "discourse translucent" (cf. also Yanovich 2008).

Opinions about anaphoric accessibility of PINs have varied widely, with little empirical evidence beyond the intuitions of the researchers or anecdotal observations. But this situation is changing. In a first experimental investigation on the processing of small discourses with bare noun antecedents in Mandarin, Law & Syrett (2017) found evidence for discourse translucency in a self-paced reading experiment. For German, Brocher et al. (2020) found evidence for a reduced prominence of PINs using eye tracking, In experiments on Persian (cf. Modarresi & Krifka 2020 and to appear a, b) that involved item selections and free sentence completion, we found clear evidence that anaphoric reference to PINs is natural, albeit slightly less straightforward than with indefinite PINs.

We take it that the current experimental evidence speaks against the opacity hypothesis for PINs, at least for the investigated types in Mandarin, German, and Persian. In the present article, we will report on additional experiments to our cited work that bear on the precise mechanism how PINs introduce DRs. We have seen that PINs allow for a number-neutral interpretation, which is in principle compatible with reference to a single object or a multitude of objects, e.g. one or more books in (1). In Modarresi & Krifka (to appear a, b) we considered experimental items in a way that should prevent object to be understood with a bias towards one or more than one object. This is arguably the case for (1), as Sara could equally likely have bought one book or more than one book. This contrasts with (2) and (3), which have a strong tendency towards a multitude vs. a single interpretation, respectively. In (2) this is due to the temporal quantification 'the whole day', in (3) this is due to stereotypical knowledge about prizes in competition.

(2) man tamam-e-ruz   livan shost.am   va baad (#oon/oona) ro/Ø   khoshk kardam.
     I   whole-Ez-day glass washed-1SG and then (#that/those)-OM/Ø dry   did.1SG
     I washed glass the whole day and then dried Ø / them.

(3) Nafar-e aval   saat-e-tala   barandeh shod   vali doos-esh/#eshoon/Ø   na-dasht
     the first winner watch-EZ-gold winner   became.3SG but like-it/#them/Ø   NEG-had.3sg
     'the first winner won golden watch but didn't like it/like'

Similarly, (4) tends to be understood as buying a multitude of carrots, whereas (5) is preferably interpreted as involving only one car. The difference is due to the stereotypical enrichment of linguistic information – carrots are typically bought in groups and cars are bought as single object.

(4) Sara havij kharid   va man poost-e-shoon/ Ø /?esh ro kandam.
     Sara carrot bought.3sg and I   skin-EZ-them/ Ø /it/ OM   peeled.1SG.
     'Sara bought carrot and I skinned them// Ø /?it'

(5) Sarah emrooz mashin kharid   va rooz-e-baad   foroukht/ Ø /esh/?eshoon
     Sara today   car   bought.3SG and day-EZ-next   sold/ø/it/#them
     'Sar bought car today and sold (Ø/it) next day'

In the current paper, we will report on experiments in which the stereotypical enrichment of the number-neutral interpretation was systematically varied. This leads to a new evaluation of the precise mechanisms by which PINs introduce DRs.

2. Two theoretical models of anaphoric uptake for PINs. As mentioned above, we will concentrate here on theories that are consistent with the finding that anaphora to PINs (more specifically, bare object nouns in Persian) is possible, following the recent experimental evidence of Modarresi & Krifka (to appear a, b). There are two ways in which such anaphora may work: directly, by the introduction of DRs by antecedent expressions that are picked up by co-referring expressions, or indirectly, by associative anaphora. Associative anaphora is illustrated in the following case:

(6) Sarah bought a book. The cover picture looked interesting.

The antecedent clause did not introduce a DR for *the cover picture*. Rather, a referent of this expression can be constructed due to the introduction of a DR for *a book*, and the stereotypical enrichment that books often have cover pictures on them. The results in the sentence continuation experiment of Modarresi & Krifka (to appear. a, b) speak against the possibility that anaphora to bare objects in Persian is predominantly by associative anaphora, as in this case we should find full DPs as the preferred case of anaphoric uptake. As full DPs were rarely produced by the participants (see also section 3.3 below), we can exclude them as a dominant way of uptake. Hence, we assume that anaphoric uptake of PINs in Persian is mediated via discourse referents.

There are different theoretical options for the uptake mediated by DRs, of which we consider two. Both theoretical options are couched in the language of Discourse Representation Theory (DRT), which we outline here to the extent that is necessary (cf. Kamp & Reyle 1993, Kamp et al. 2011 for comprehensive introductions). DRT assumes a semantic representation in terms of discourse representation structures (DRSs), which are pairs of a set of accessible DRs and conditions on these DRs. DRSs are typically depicted in box format but we will render them here more compactly as pairs of the form ⟨discourse referents | conditions⟩. A sentence is interpreted as

expanding the DRS of the preceding discourse, where non-anaphoric DPs introduce new DRs, and anaphoric DPs pick up DRs that were already introduced. This can be illustrated for *yek*-marked as follows:

(7) a. Sara yek ketāb kharid.　　　　b. Ø/Oo　foran　　　khoond-esh.
　　　　Sara　one book bought　　　　　　　(s)he immediately read-it
　　　　'Sarah bought a book.'　　　　　　　　'She/he immediately read it.'

(8) $\langle \, | \, \rangle + (5)(a) = \langle x_1\ x_2\ e_1 \mid x_1 = $ Sara, book(x_2), $|x_2| = 1$, e_1: bought$(x_1,x_2)\rangle$

(9) $(13) + (7)(b) = \left\langle \begin{array}{c|c} x_1\ x_2\ e_1 & x_1 = \text{Sara, book}(x_2), |x_2| = 1, e_1\text{: bought}(x_1,x_2) \\ e_2 & \text{immediately\_after}(e_1,e_2), e_2\text{: read}(x_1,x_2) \end{array} \right\rangle$

Here, (8) describes the update of an empty initial DRS $\langle \, | \, \rangle$ by the first sentence, which introduces a DR x_1 for Sara, a DR x_2 for one book, and a DR e_1 for a past event of buying of x_2 by x_1. In the conditions, we use the format of Kamp & Reyle (1993); in particular, $|x|$ specifies the number of atomic entities that the DR x is anchored to, and e: R(x, y) states that e is an event in which x and y stand in the relation R to each other. The resulting DRS is further expanded in (9) by the second clause, which picks up x_1 and x_2 by the anaphoric devices of an subject pronoun realized as *oo* or empty, as Persian is a pro-drop language, and an object clitic *-esh*. The second clause also introduces another event DR e_2 that is immediately after e_1 and is a past reading of x_2 by x_1.

　　DRSs are interpreted with respect to a model M that contains a set of entities that have certain properties and stand in certain relations to each other. A DRS $\langle \, D \, | \, C \, \rangle$ with a set of DRs D and a set of conditions C is true with respect to a model M iff there is a function g that maps all DRs in the set D to entities in M such that all conditions in C are true for the corresponding entities in M.

　　The first theory involving DRs was proposed by Modarresi (2014, 2015). The *yek*-marked singular indefinite object in (7) introduces a DR that is anchored to a single book (cf. the condition $|x_2| = 1$). Modarresi assumes that bare objects differ minimally insofar as they introduce number neutral DRs (already assumed by Kamp & Reyle 1993 for different phenomena), which are given here by Greek letters ξ. This is justified by the number-neutral interpretation of bare objects in Persian (and PINs in general).

(10) a. Sara ketāb kharid.　　　　　　　b. Ø/(Oo)　khoond-Ø/-esh/-eshoon.
　　　　Sara book bought　　　　　　　　　(s)he　　read-Ø/-it/-them
　　　　'Sarah bought a book.'　　　　　　　　'She/he read it.'

(11) $\langle \, | \, \rangle + (7)(a) = \langle \, x_1\ \xi_2\ e_1 | \ x_1 = $ Sara, book(ξ_2), $|\xi_2| \geq 1$, e_1: bought$(x_1,\xi_2) \, \rangle$

(12) $(13) + (7)(b) = \left\langle \begin{array}{c|c} \xi_1\ x_2\ e_1 & x_1 = \text{Sara, book}(\xi_2), |\xi_2| \geq 1, e_1\text{: bought}(x_1,\xi_2) \\ e_2 & e_2\text{: read}(x_1,\xi_2), |\xi_2| \geq /= />1 \end{array} \right\rangle$

We represent the fact that number-neutral DRs ξ can be anchored to atomic entities or to sum individuals consisting of more than just one atomic entity by the condition $|\xi| \geq 1$. In this case, the anaphoric uptake is natural with a null anaphor, which does not restrict the DR to any particular number. But uptake is also possible by the singular enclitic anaphor *-esh* and the plural enclitic anaphor *-eshoon*. In the latter cases, the anaphoric expression contains additional information. Such "specificational" anaphora are known for gender (e.g. *How do you think God looks like?* – *Well, I think she is black*, where the pronoun *she* resolves the underspecified gender of the antecedent *God* to female).

Farkas & de Swart (2003), discussing Hungarian, claim that PIN objects can be picked up by null anaphora. Modarresi (2014, 2015) assumes that this is the preferred uptake in Persian as well and explains this by null anaphora not having a number feature. But Modarresi also allows for singular anaphora (-*esh*) and plural anaphora (-*eshoon*), as their semantic restrictions are compatible with the number neutrality of PIN antecedents. In these cases the anaphoric expression are specificational, as they carry additional information (that Sara bought one book, or that Sara bought more than one book). Such additional information may be supported by stereotypical interpretations, as in (4) and (5) for the plural and singular interpretation, respectively. Hence, the preferred stereotypical interpretation should have an influence on the nature of the anaphoric uptake of PINs. According to Modarresi (2014, 2015), there is no fundamental difference in the anaphoric potential of *yek*-marked objects and bare objects in Persian; they both introduce a DR that is accessible for future uptake. Modarresi does not assume that singular, plural or neutral DRs differ in their markedness; if there are differences, we should assume that number-neutral DRs are least marked.

The second theory we consider is Krifka & Modarresi (2016). According to it, the event DR is bound by a narrow-scope existential quantifier, the existential closure introduced by Diesing (1992), which scopes over the syntactic domain of the vP. Objects with a *rā* scramble out of the vP -hence escape existential closure- and have to be interpreted outside of the scope of the existential quantifier (Modarresi 2014). *Yek*-marked objects not marked by *rā* stay within the vP but can scope within or outside of existential closure, a variability that is known for indefinites with determiners in general (cf. Fodor & Sag 1982). A further assumption that sounds unintuitive initially is that bare nouns are definites, with a singular interpretation. This holds uncontroversially for subjects and for objects marked by *rā*, which tend to have a definite, number-specific interpretation, cf. (1). But we take bare objects not marked by *rā*, which remain within the vP, to be singular definites as well. This is possible because we assume that bare nouns in general are dependent definites. The apparent indefinite number-neutral interpretation of bare objects without *rā* marking arises as a secondary effect due to the place where the bare noun is interpreted, within the scope of the existential quantifier over the event, and as functionally related to the event. One theoretical advantage of his hypothesis is that a uniform interpretation of bare nouns as singular definites as subject, *rā*-marked objects, and objects that are not *rā*-marked becomes possible.

The narrow-scope indefinite, number-neutral interpretation of bare objects comes about as illustrated in the following examples. The wide-scope interpretation of example (7) with *yek*-marked object is given in (13) and (14):

(13) $\langle \ | \ \rangle + (7)(a) = \langle x_1 \ x_2 \ | x_1 = \text{Sara}, \exists \langle e_1 | \ \text{book}(x_2), |x_2| = 1, e_1: \text{bought}(x_1, x_2) \rangle \rangle$

(14) $(13) + (7)(b) = \left\langle x_1 \ x_2 \ \left| \begin{array}{l} x_1 = \text{Sara}, \exists \langle e_1 | \ \text{book}(x_2), |x_2| = 1, e_1: \text{bought}(x_1, x_2) \rangle \\ \exists \langle e_2 | \ e_2: \text{read}(x_1, x_2) \rangle \end{array} \right. \right\rangle$

Notice the condition of the form $\exists \langle \ D \ | \ C \rangle$, where $\exists$ stands for Diesing's existential closure operator. This is a complex condition, other examples of complex conditions being negation, disjunction, and quantification (cf. Kamp & Reyle 1993). The condition $\exists \langle \ D \ | \ C \rangle$ holds with respect to a function g and a model M iff g can be extended to a function g′ that also maps the DRs in D to entities in M such that all the conditions in C are satisfied in M. The resulting DRS (14) is truth-conditionally equivalent to (9).

The interpretation of (10) is given as follows, on the input of an empty DRS $\langle \ | \ \rangle$.

(15) $\langle \ | \ \rangle + (10)(a) = \langle x_1 \ | \ x_1 = \text{Sara}, \exists \langle e_1 \ x_2 | \ x_2 = \text{book-of}(e_1)(x_2), |x_2| = 1, e_1: \text{bought}(x_1, x_2) \rangle \rangle$

Here the object is interpreted as a definite dependent on the event e_1, as 'the unique (single) book of e_1'. As e_1 is introduced within existential closure, the corresponding DR x_2 has to be introduced within existential closure as well. A consequence of this is that x_2 cannot be accessed directly by the following sentence. But Kamp & Reyle (1993) have proposed, for quite different reasons, that DRs in subordinated DRSs can be reactivated by an operation called abstraction and summation. For the case at hand, this is illustrated in (16).

$$(16) \quad (15) + (10)(b) = \left\langle \begin{array}{c|c} x_1\ x_2 & x_1 = \text{Sara}, \exists\langle e_1\ x_2 |\ x_2 = \text{book-of}(e_1)(x_2), |x_2| = 1, e_1: \text{bought}(x_1,x_2)\rangle \\ x_3 & x_3\ =\ \Sigma x_2\ \langle e_1\ x_2|\ x_2 = \text{book-of}(e_1)(x_2), |x_2| = 1, e_1: \text{bought}(x_1,x_2)\rangle \\ & \exists\langle e_2|\ e_2: \text{read}(x_1,x_3)\rangle, |x_3| \geq/=/> 1 \end{array} \right\rangle$$

In the second line, a new DR x_3 is introduced and anchored to the sum (Σ) of all x_2 that satisfy the condition expressed in the scope of Σ, namely that there is an event e_1 such that x_2 is the unique book of e_1 and e_1 is an event of x_1 buying x_2. This is the sum of all books that Sara bought, in the relevant discourse universe. Notice that this sum can be one or more than one book. The DR x_3 can be taken up in the second sentence by a DR that is number-neutral ($|x_3| \geq 1$) as with null anaphora, or atomic ($|x_3|=1$) as with the singular anapher -*esh*, or non-atomic ($|x_3|>1$) as with the plural anaphor -*eshoon*. In contrast, in case of a *yek*-marked indefinite as in (14), only a neutral or singular anaphor is possible.

This approach differs from the assumption of number-neutral DRs, as it assumes a more complex mechanism of anaphoric update in the case of bare noun objects compared to *yek*-DPs. In general, we should find that anaphoric uptake for bare noun (PIN) is less frequent than with *yek*-marked nouns. This is what Modarresi & Krifka (to appear a, b) indeed found, in particular in their free sentence completion task (see section 3.3 below).

Notice that just as for the approach with number-neutral DRs in Modarresi (2014, 2015), cf. (12), this analysis predicts an influence of stereotypical world knowledge on the use of singular or plural anaphoric devices. If world knowledge suggests that more than one entity was subjected to the event, as in (4), we should easily find the plural anaphor -*eshoon* next to the null anaphor, but the singular anaphor -*esh* should not occur. This is in contrast to cases like (5) which suggest that only one entity is involved.

But the analysis presented here differs from Modarresi (2014, 2015) in one respect. The simplest summation is in case there is only one relevant event in the model, as it then amounts to referring to the single atomic individual that is involved in that event. In this limiting case, the summation operation Σ is reduced to the identifying function $x_3 = \iota x_2 \langle e_1\ x_2 |\ \dots \rangle$. Hence, we should find, in addition to an effect of stereotypical world knowledge, a preference for anaphoric uptake by singular pronouns, instead of null or plural pronouns, as simple summation would yield DRs that are anchored to atomic individuals. This is different to Modarresi (2014, 2015), who assumes number-neutral DRs; in this case, number-neutral null anaphora should be the preferred choice of anaphoric uptake.

3. Experimental evidence. In this section we will present three experiments that provide evidence for the hypotheses presented in the previous sections. In order to facilitate the discussion, we list here five hypotheses, where A and A* are related to Modarresi (2014, 2015) according to which PINs are discourse transparent, and B and B* are related to Krifka & Modarresi (2016) according to which PINs are discourse translucent. 0 is the hypothesis that PINs do not introduce DRs, hence that they are discourse opaque.

(17) Specific hypotheses:
   0:    Bare noun (BN) objects do not license anaphora.
   A: BN objects license anaphora to the same degree as *yek*-marked (YK) objects.
   B: BN objects license anaphora to a reduced degree compared to YK objects.
   A*: BN objects introduce number-neutral DRs,
          preferred uptake by null anaphora
   B*: BN objects introduce DRs by existential closure,
          preferred uptake by singular anaphora

Experiment 1 involves forced choice of anaphora; Experiment 2 forced choice of antecedents; Experiment 3 is a free text completion task. Hence, all three experiments test language production.

3.1. FORCED CHOICE OF ANAPHORA. In this experiment we tested the anaphoric potential of bare noun (BN) objects and *yek*-marked (YK) objects in a forced-choice selection of anaphoric expressions, a controlled production experiment. Participants were presented with a sentence containing a BN object or YK object in antecedent sentences that were constructed in a way as to have a bias towards a singular or, a plural interpretation, or no particular interpretation (neutral bias).

   The continuation sentence contained a blank to be filled by null, singular (SG), or plural (PL) anaphora. The reason for testing three types of biases was based on the observation by Modarresi (2014) that such biases may affect the choice of pronominal anaphora referring to BN antecedent. We constructed 36 test items including 8 fillers. The test items had 6 conditions: (2 antecedent types and 3 bias types). As a sample item representing all six conditions and the possible three reactions, consider (18). The experimental items were presented in Persian script, of course.

(18) Sara { yek / --- } {television / ketāb / havij } kharid.
      Sara   IDF / BN   TV   book   carrot bought

      Baad tu-ye mashin   ☐ gozasht-esh   ☐ gozasht   ☐ gozasht-eshoon
      then in-EZ car   put-it   put-Ø   put-them

We list the test items in shortened form for singular bias (19), neutral bias (20) and plural bias (21) in Persian together with an idiomatic translation. We assigned the bias category following our own intuition, also asking other native Persian speakers. While it is relatively easy to find examples with clear singular or plural bias, the construction of examples with neutral bias are less clear-cut.[1]

(19) *khooneh be ers bord* 'inherited house', *mashin lebasshoui kharid* 'bought washing machine', *motorcyclet kharidam* 'bought motorcycle', *lebase aroos keraye kard* 'rented wedding dress', *baraye doostash sandwich kharid*, 'bought sandwich for a friend', *baraye behtarin danesh-amooz jayezeh sefaresh dad* 'ordered prize for the best student', *docharkhe did* 'saw bicycle', *keike tavallod avard* 'brought birthday cake', *ghayegh ejareh kard* 'rented boat', *gavsandogh dozdid*, 'stole safe', *khooneye jadedemoon piano dareh* 'our new house has piano', *saate tala barandeh shod* 'won golden watch'

---

[1] The construction of examples could have been objectivized based on a very large corpus. For example, Google N-grams shows that the rate of occurrences of the string *had a piano* vs. *had pianos* is about 14, of *received a gift* vs. *received gifts* is about 1.12, and of *corrected a paper* vs. *corrected papers* is about 0.01, indicating a clear single, neutral, and plural bias, respectively. This is based on English; no comparable large corpus for Persian is available.

(20) *jabeh avord* 'brought box', *ketāb kharid* 'bought book', *sham roshan-kard* 'blew candle', *abnabat avord* 'brought bonbon', *ketāb khoond* 'read book', *card-postal ferestad* 'sent postcard', *livan kharid* 'bought glass', *hedyeh gereftam* 'received gift', *badkonak kharid* 'bought balloon', *ketāb gharz gereft* 'borrowed book', *too ashpazkhooneh magas did* 'saw mosquito in the kitchen', *az ghafaseh fenjoon biroon avard* 'took cup from the shelf'

(21) *adas rikht* 'poured lentil', *zardaloo chidam* 'picked apricot', *baghboonha too-ye-bagh derakht kashtand* 'the gardeners planted trees in the garden', *too dasht shaghayegh daroomadeh* 'in the pasture bloomed tulip', *baraye sakhtane divar ajor sefaresh-dad* 'ordered brick to build wall', *sherkate rahahan barayaerahhaye mokhtalef ghatar kharid* 'the railroad bought train for different roads', *baraye mehmooni sandali sefaresh dadam* 'ordered chairs for garden party', *moallem baraye bachehha ye class jayezeh gereft* 'the teacher got prize for children in the class', *too bazar havij mifrookht* 'sold carrot in bazar', *tamame rooz livan shostam* 'washed glass all day', *tamame rooz varagheh sahih kard* 'corrected paper all day', *tamame hafteh daman dookht* 'sew skirt all week'

There were 357 native Persian speakers that voluntarily participated in this experiment using an online survey platform (Socsi survey). The stimuli were presented in twelve different lists.[2] Each list included all six conditions, with an average of four fillers; the items were randomized using Latin square design. The results are indicated in Figure 1.
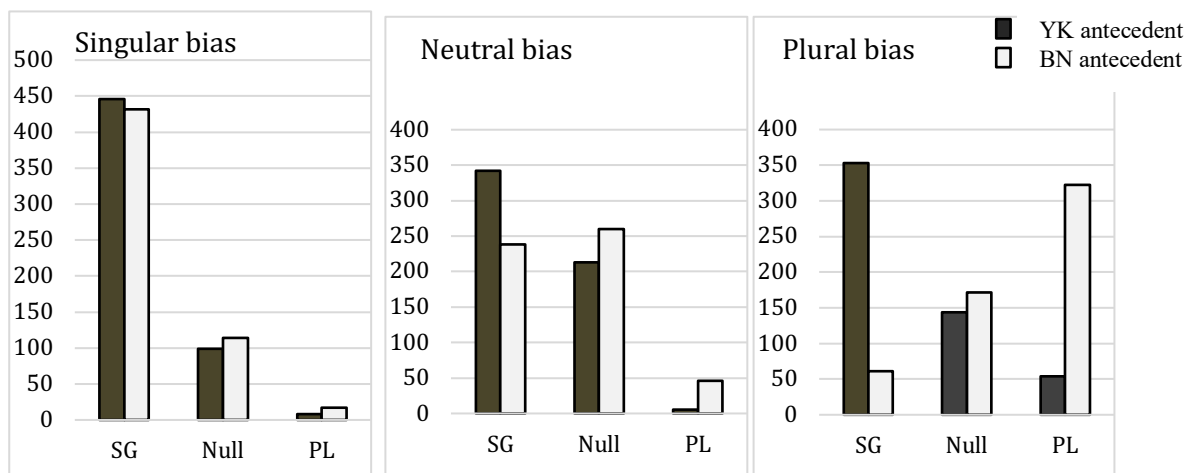

Figure 1: Forced choice of SG, null and plural anaphora in sentences with singular, neutral or plural bias and *yek*-marked antecedent or BN antecedent.
Y-axis specifies number of items.

As participants were forced to select an anaphoric uptake, the experiment cannot distinguish between hypotheses 0 and A/B. But it showed that bias has an effect on the nature of the anaphoric uptake. BN antecedents were taken up most often by SG pronouns under singular bias, and by PL pronouns under plural bias, consistent with hypotheses A* and B*. Comparing YK and BN

---

[2] Items from the next experiment, antecedent choice, were also included, that is why we randomized in twelve lists as opposed to six lists. It was made sure that no participant saw the same sentence twice.

antecedents, we find that BN antecedents favored Null anaphora a bit more under all three bias conditions. This can be interpreted as a slight preference for hypothesis A* over B*. However, we would expect a considerably stronger preference if Null anaphora corresponds to BN antecedents in their number feature, as assumed under hypothesis A*. Furthermore, under the Neutral bias condition, BN antecedents are taken up by SG and Null anaphors equally. This would be expected under hypothesis B*, which assumes a structural tendency for simple summation, in contrast to A*.

3.2. FORCED CHOICE OF ANTECEDENT. Experiment 1 did not show whether the participants favored the use of BNs as antecedents of anaphora because the antecedents were fixed conditions in the experimental items. We reversed the design and investigated the choice of antecedents (BN vs. YK objects), when the anaphor in the subsequent sentence is fixed (as NL, SG or PL). Like in the previous experiments we had three types of biases. With the exception of the reversal of the design, the stimuli were the same as in Experiment 1. This is illustrated in (22).

(22)   Ali □ { television / ketāb / havij} / □ yek { television/ ketāb / havij}   kharid va
       Ali   TV   book carrot}   IDF TV   book/ carrot} bought   and

       { gozasht-esh /   gozasht-Ø   / gozasht-eshoon }   rooy-e-miz.
         put-it   put-Ø   put-them   on-EZ-table

There were 36 items with 9 conditions (3 anaphor types x 3 bias types), including 8 fillers. The same 357 native Persian-speakers as in Experiment 1 participated in Experiment 2, as the second part of the experiment. The stimuli were presented in twelve different lists. Each list included all the nine conditions in randomized order, one of the conditions of each sentence, including an average of four fillers in each list, using a Latin square design. After reading the whole sentence, the participants had select the BN or the YK noun as the most appropriate antecedent. Results are presented in Figure 2.
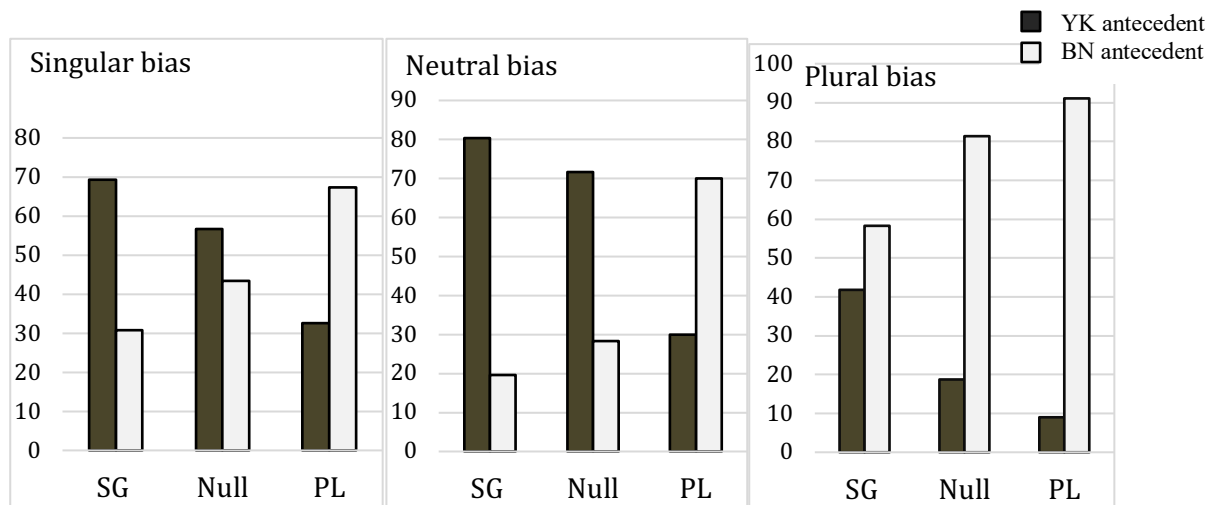

Figure 2: Forced choice of *yek*-marked antecedent of BN antecedent with SG, null and PL anaphora and singular, neutral and plural bias. Y-axis specifies number of items.

We concentrate first on the cases with singular and neutral bias. Clearly, YK objects make better antecedents except for PL anaphors, as in this case there would be a semantic clash between

the singular-marked antecedent. This supports hypothesis B over A (the hypothesis that BN and YK-marked antecedents have the same frequency for singular/neutral bias and SG/Null anaphors is rejected by a chisquare test with $p < 0.001$).

But the experiment also shows that BN objects were actually selected quite often, even for SG and Null anaphors. This is evidence against hypothesis 0. BN objects are naturally favored in cases with plural bias, which disfavors YK antecedents for semantic reasons. Notice that these semantic reasons outweigh any preference of YK antecedents as anaphors, even in the case of SG pronouns.

Surprisingly, even with PL anaphors there was a substantial minority of cases with YK antecedents (about 30 vs. 70 for the cases with singular and neutral bias). A possible reason is that the YK objects were interpreted with narrow scope, introducing their DR in the existentially quantified sub-DRS, which would be consistent with hypothesis B*. An alternative explanation is that the task of going back in the text, choosing an antecedent, reading the text, choosing the other antecedent, reading the text in this version, and selecting the better option of the two versions was quite complex. It might have led to selecting the YK variant without reading the whole sentene, because this is in general the better antecedent.

3.3. FREE COMPLETION TASK. In a final experiment, we investigated which anaphoric forms are generated spontaneously in a free completion of a preceding sentence, contrasting BN and YK-marked antecedents. This task does not investigate the reflection of participants about language but rather asks for a natural production task, leaving many more options. In particular, it also leaves open the option of no anaphoric uptake at all. A sample item is (23).

(23)   Leila { yek / --- } { television / ketāb / havij }   kharid   va baad _____
        Leila {IDF / BN }{ TV   book   carrot } bought.3SG and then
        'Leila bought (a) TV/ book / carrot and then _____ ,'

There were 6 experimental conditions (2 antecedent types x 3 bias types). The stimuli consisted of 36 items including 3 fillers, randomized in a Latin Square design in eight lists. There were altogether 252 participants that took part in an online experiment. Participants read sentences in different conditions and were asked to type a suitable continuation. We collected about 330 to 420 data points for each condition, altogether 2256 data points 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 3 visualize NL anaphora, singular anaphoric reference with pronouns or clitics (Pro-SING), singular anaphoric reference will 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.[3]

---

[3] 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, presumably 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. For this reason, associative anaphora of the type *Leili got married. He is nice.* are impossible in Persian.
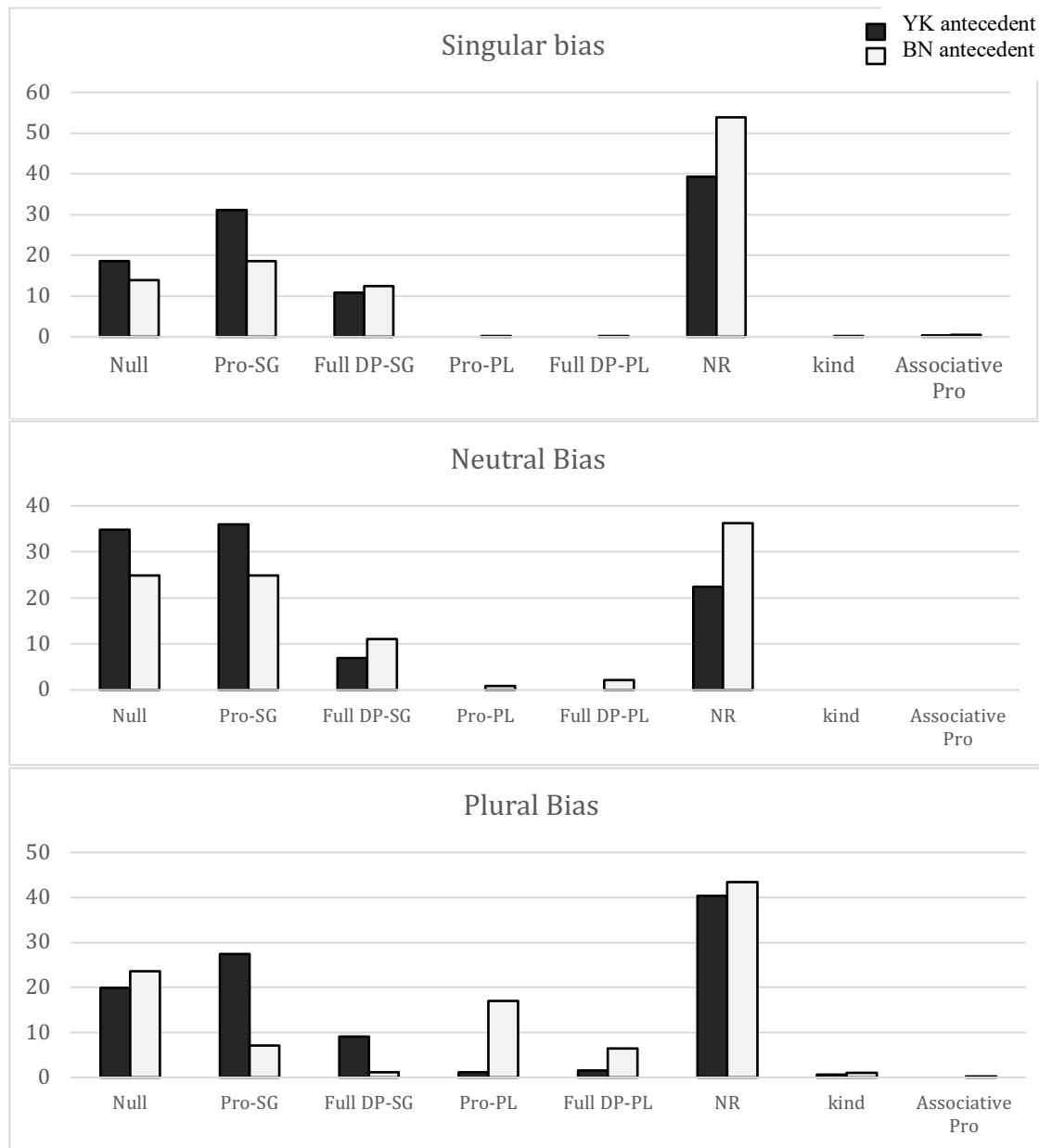
Figure 3: Free sentence completion task with YK vs. BN antecedents in sentences with singular, neutral and plural bias. Pronominal uptake by Null anaphora, SG pronouns, full singular DPs, plural pronouns, full plural DPs, no reference, kind reference, and associative pro-forms. Y-axis specifies percentages.

We discuss here the main results from the completion experiment. First, we see in the NR column that BN objects are picked up about half of the time (slightly less so in the singular bias). This is definite evidence against hypothesis 0. Uptake is only in a minority of cases by full DPs, making it implausible that the uptake is predominantly by associative anaphora. Concentrating on the singular and neutral bias situations, we see in the NR column that BN objects are less often picked up by anaphora than YK objects, supporting hypothesis B over A. We also see that BN antecedents do not particularly favor Null anaphors, supporting hypothesis B* over A*. Uptake of BN antecedents by Null and singular anaphora is about equal, arguing against hypothesis A*,

which predicts a greater affinity of BN antecedents to null anaphora. This is consistent with hypothesis B*, which states that there is a tendency to prefer simple summation, licensing singular anaphora. One result that is difficult to interpret is why there is more anaphoric uptake in the cases of neutral bias, compared to singular and plural bias.

4. Conclusion. In this article we discussed the nature of pseudo noun incorporation (PIN), with Persian bare noun objects (BN) as example. Most theories discuss semantic properties of pseudo incorporation, which includes their unspecificity as to singular or plural interpretation. We argue that the anaphoric potential, the ability to be taken up by anaphoric expressions, is crucial for the proper analysis of PINs. We investigated their anaphoric potential in contrast to regular singular indefinites marked with the determiner *yek* 'a / one' in experimental items with three types of biases: bias towards a singular interpretation, towards a plural interpretation, and with neutral bias that neither favors singular nor plural interpretation.

Despite the widespread assumption that PINs do not introduce discourse referents (DRs) (for Persian BNs, cf. Modarresi & Krifka to appear a), our experimental results have shown that bare nouns are actually quite good antecedents – though slightly less than indefinite antecedents.

The focus of the current paper was on the effect of biases towards singular or plural interpretations of PIN objects, and the absence of such biases. The experiments provided evidence that can be cautiously interpreted as disfavouring theories that assume that PIN objects are semantically specified as number-neutral such as Modarresi (2014, 2015), as then we would have found a greater preference for number-neutral null pronouns for PIN objects. Rather, the experiments favour the proposal by Krifka & Modarresi (2016), which considers PIN objects as dependent singular definites within existential closure over an event variable. The DR of PINs is not directly available but can be accessed via a process of abstraction and summation, a phenomenon that is well-known in other cases, as in so-called donkey sentences. The process predicts a general preference for singular DRs, for which there is evidence in our experimental data.

References

Aguilar-Geuvara, Ana & Zwarts, Joost. 2010. Weak definites and reference to kinds. *Semantics and Linguistic Theory (SALT)* 20, 179–196.

Borik, Olga & Berit Gehrke. 2015. An introduction of the syntax and semantics of pseudo-incorporation. In: Borik, Olga & Berit Gehrke (eds.), *The syntax and semantics of pseudo-incorporation*. Leiden: Brill, 1-46.

Brocher, Andreas et al. 2020. Referent management in discourse: The accessibility of weak definites. *Cognitive Science* 2829-2835.

Chung, Sandra & William Ladusaw. 2020. Noun incorporation. In Gutmann, Daniel et al. (eds), *The William Blackwell Company to Semantics*.

Dayal, Veneeta. 2011. Hindi pseudo-incorporation. *Natural Language and Linguistic Theory* 29, 123–167.

Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.

Dobrovie-Sorin, Carmen & Ion Giurgea. 2015. Weak reference and property denotation: Two types of pseudo-incorporated nominals. In: Gehrke, Berit & Olga Borik, (eds), *The syntax and semantics of pseudo-incorporation*. Brill, 88-125.

Espinal, M. Teresa & Louise McNally. 2011. Bare nominals and incorporating verbs in Spanish and Catalan. *Journal of Linguistics* 47, 87-128.

Farkas, Donka, and Henriëtte de Swart. 2003. *The semantics of incorporation: From argument structure to discourse transparency*. Stanford: CSLI Publications.

Fodor, Janet Dean & Ivan A. Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5: 355-398.

Ghomeshi, Jila. 2008. Markedness and bare nouns in Persian. In: Simin Karimi, Vida Samiian & Donald Stilo (eds), *Aspects of Iranian Linguistics*. Cambridge Scholar Publishing. 85-112.

Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic. Introduction to model theoretic semantics of natural language, formal logic, and Discourse Representation Theory.* Dordrecht: Kluwer.

Kamp, Hans, Uwe Reyle & Josef Van Genabith. 2011. Discourse Representation Theory. In: Guenthner, Franz & Dov M. Gabbay, (eds), *Handbook of Philosophical Logic*. Springer, 125-394.

Krifka, Manfred & Fereshteh Modarresi. 2016. Number neutrality and anaphoric uptake of pseudo-incorporated nominals in Persian (and weak definites in English). *SALT* 26, 874-891.

Ladusaw, William & Sandra Chung. 2004. *Restriction and saturation*. Cambridge, MA: MIT Press.

Law, Jess H.-K. & Kristen Syrett. 2017. Experimental evidence for the discourse potential of Mandarin. *North Eastern Linguistics Society (NELS)* 47.

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language and Linguistic Theory* 19: 153-197.

Massam, Diane. 2017. Incorporation and pseudo-incorporation in syntax. In: Aronoff, Mark, (ed), *Oxford Research Encyclopedia of Linguistics*.

McNally, Louise. 1995. Bare plurals in Spanish are interpreted as properties. In Formal Grammar, eds. Glyn Morrill & Richard Oehrle (eds.), *Formal Grammar*. Barcelona: Polytechnic University of Catalonia, 197-212.

McNally, Louise & Veerle van Geenhoven. 2005. On the property analysis of opaque complements. *Lingua* 885-914.

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 Olga Borik & Berit Gehrke (eds.), *The syntax and semantics of pseudo-incorporation*. Leiden: Brill, 189-221.

Modarresi, Fereshteh & Manfred Krifka,. 2020. Anaphoric potential of pseudo incorporated nominals in comparison with compounds and implicit objects. *Linguistic Evidence* 2020.

Modarresi, Fereshteh & Manfred Krifka. To appear a. Anaphoric potential of pseudo-incorporated bare objects in Persian. In Simin Karimi (ed), *North American Conference on Iranian Languages* (NACIL2).

Modarresi, Fereshteh & Manfred Krifka. To appear b. Pseudo-incorporation and anaphoricity: Evidence from Persian. *Glossa*.

Schwarz, Florian 2014. How weak and how definite are weak indefinites? In: Aguilar-Guevara et al. (eds), *Weak referentiality*. Amsterdam: John Benjamins.

van Geenhoven, Veerle. 1998. *Semantic incorporation and indefinite description*. Standford: CSLI Publications.

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, 367-379.