In Deniz Zeyrek & Umut Özge (eds.), Discourse meaning: The view from Turkish, 81–102. Berlin: De Gruyter.

Umut Özge and Klaus von Heusinger

Case marking and forward and backward discourse function

Abstract: The paper reports a corpus search and annotation study investigating the discourse functions of Differential Object Marking (DOM for short) in Turkish, which is manifested as optional accusative case on indefinite direct objects. Turkish DOM has been associated with specificity, presuppositionality and wide scope behavior with respect to other sentence operators. This sentence semantics has been related to different properties of discourse prominence: First to a backward discourse-linking function, where DOM-marked indefinites are indicated to have an antecedent in the discourse, without, however, there being a consensus on the type of the relation between the antecedent and the indefinite. Second, DOM in Turkish is also assumed to be related to forward direction in discourse under the general notion of being more likely to be talked about in the ensuing discourse. In order to test whether we can assign these two functions to case marking, we searched DOM tokens from a 21M corpus and annotated them with respect to categories relevant for both backward and forward discourse functions. Contrary to previous proposals and our assumptions regarding backward and forward linking, we did not observe any discourse function in either direction in our data set. We provide some arguments for why Turkish case marking does not show the discourse functions that are demonstrated for other languages.

1 Introduction

Indefinite noun phrases are the "standard" linguistic devices for introducing new referents to a discourse model. They show different referential strength in a sentence. They all express an existential entailment and introduce a discourse referent. However, the properties of the discourse referent do vary according to the interpretation of the indefinite: A specific indefinite shows wide scope and introduces a stable discourse referent (Kamp and Bende-Farkas 2018), a regular

Umut Özge, Cognitive Science Department, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey
Klaus von Heusinger, Department of German Language and Literauture 1, University of Cologne, Cologne, Germany

https://doi.org/10.1515/9783110686654-004

indefinite shows scope interactions with other operators, narrow scope with respect to scope islands and introduces short term discourse referents in the scope of negation or intensional operators. Weak indefinites (pseudo-incorporated indefinites) always display narrow scope and may not introduce a discourse referent (Kamp 2014). In this paper we focus on specific indefinites in Turkish. Specificity is often associated with discourse prominence – we focus on two properties of discourse prominence: backward looking functions such as d-linking or partitivity (Enç 1991), and forward looking functions, such as topic shift, referential persistence (Chiriacescu and von Heusinger 2010), noteworthiness (Ionin 2006) or cataphoricity (Deichsel and von Heusinger 2011). First, there are various ways an indefinite can be linked to the previous context. Von Heusinger and Özge (submitted) distinguish two kinds of backward linked indefinites: inferentially linked indefinites and partitives. Prince (1981, 1992) has observed that certain types of indefinites are inferentially linked to some components of the discourse model they get introduced into. As Prince (1992: 306, ex. 19) observes, while the page reported to fall out of the book in (1b) is not *any* page, but one that belongs to the book under discussion, the same does not hold for the cockroach, showing that the indefinite *a page* is inferentially anchored to an entity already established in the discourse.

(1) a. I picked up that book I bought and *a cockroach* fell out.
    b. I picked up that book I bought and *a page* fell out.

Enç (1991) discusses a different kind of discourse linking, namely partitivity, as in (2), where the indefinite *two girls* in (2b) has either a partitive or a non-partitive reading. In the partitive – and according to Enç (1991) specific – reading, the two girls belong to the set of children introduced in (2a). In the non-partitive reading, the girls are either not part of the set of children introduced in (2a) or the speaker is not intending to make any claim about this.

(2) a. Some children entered the room.
    b. I know two girls.

Specific or strong indefinites are also associated with forward linking, i.e. with signaling that the referent they introduce will be taken up in the upcoming discourse. A straightforward example of this phenomenon is the use of English indefinite *this*. Ionin (2006: 180; quoting MacLaran 1982: 88) demonstrates in (3) that the use of indefinite *this* in English is only licensed if the associated referent has a noteworthy property and is re-used, while the regular indefinite article does not show this property.

(3) a. I put **a/*this** 1$ stamp on the letter. I wanted to mail the letter to Europe.
b. I put **a/this** 1$ stamp on the letter and realized too late, that it was worth a fortune.

For another instance, Chiriacescu and von Heusinger (2010) show that *pe* marked indefinites in Romanian – another instance of DOM – behave like English indefinites headed by *this* in becoming more prominent in the ensuing discourse, again in comparison to non-marked indefinites.

The aim of the present paper is to investigate these two discourse functions of Differential Object Marking (henceforth DOM) in Turkish. We evaluate two hypotheses regarding the discourse functions of DOM, which are based on research on DOM in Turkish or other DOM languages.

H1 Enç (1991): there is a bidirectional implication between DOM and discourse linking (or (implicit) partitivity). We refer to this as "backward-linking" function.

H2 DOM increases the likelihood of the speaker to continue talking about the referent introduced via the indefinite in question. We refer to this as "forward-linking" function.

Turkish seems to be the optimal testing ground for these two hypotheses since it brings two phenomena of interest together. One is the discourse properties of indefinite noun phrases. Indefinites are important for discourse as they are the primary means of introducing referents, as we mentioned above. The other is case, which is also of great theoretical interest in the context of discourse studies. Case is usually thought to be a grammatical device relevant for sentence level semantics, namely in indexing the thematic arguments in a verbal constellation. However, it is shown that semantic/pragmatic contribution of case reaches beyond the sentence. Turkish provides a nice ground for studying these two phenomena, indefinites and case, since accusative marking is optional for indefinite objects in Turkish.

To address these questions concerning the discourse functions of DOM in Turkish, we conducted a corpus search and annotation study with a 21M word corpus of news texts. We annotated and analyzed the properties of the indefinites retrieved from the corpus with respect to how they relate to the preceding and succeeding discourse.

The paper is organized as follows. Section 2 provides a brief background on Turkish and DOM in this language; Section 3 gives the background for the corpus used in the study and our search and annotation pipeline; Section 4 and Section 5

reports and discusses our findings in the backward and forward directions in discourse, respectively; Section 6 concludes the paper.

2 Turkish and DOM

Turkish is a head-final language with dominant SOV clause order and rich case and verbal morphology. The nominal paradigm of Turkish is as follows:[1]

(4) a. *Ali* *bu mektub-u* *yazdı.* *Demonstrative*
A. this letter-ACC wrote.3SG
'Ali wrote this letter'

b. *Ali* *mektub-u* *yazdı.* *Definite*
A. letter-ACC wrote.3SG
'Ali wrote the letter.'

c. *Ali* *bir mektub-u* *yazdı.* *specific*
A. a letter-ACC wrote.3SG
'Ali wrote a certain letter.'

d. *Ali* *bir mektup* *yazdı.* *indefinite*
A. a letter wrote.3SG
'Ali wrote a letter.'

e. *Ali* *mektup* *yazdı.* *bare*
A. letter wrote.3SG
'Ali did some letter writing.'

Accusative case marking is obligatory with demonstrative and definite noun phrases (4a-b), while being optional with indefinite noun phrases, i.e. noun phrases preceded by the indefinite article *bir* (4c-d), and absent in bare nouns (4e). The use of the accusative marker on the indefinite direct object (see 4c versus 4d) is optional in the context of many but not all Turkish verbs, manifesting a case of Differential Object Marking (or DOM for short). Note that case-marked direct objects without the indefinite article *bir* are interpreted as definites, cf. (4b), while the indefinite article signals an indefinite reading. We further assume that unmarked direct objects without a determiner, cf. (4e), are very weakly indefinite

1 The following abbreviations are used in the examples: 3SG: 3rd person singular; ACC: accusative; DAT: dative; GEN: genitive; INF: infinitive; POSS: possessive; PROG: progressive; PST: past.

or even incorporated (see Seidel 2019 for a comprehensive discussion). We focus on the contrast between (4c and 4d).

The interpretative effects of the presence versus absence of the marker has received considerable attention in the literature (see von Heusinger and Kornfilt 2005; Özge 2011, for reviews), and can be summarized as expressing different types of specificity, such as epistemic specificity in (5), scopal specificity in (6) and referential specificity in (7) (see Fodor and Sag 1982, and for a comprehensive overview on types of specificity: von Heusinger 2011, 2019):

(5) *Mustafa bir sandalye(-yi) al-dı.*
Mustafa a chair(-ACC) buy-PST.3SG
'Mustafa bought a chair.'

(6) *Her oyuncu bir kostüm(-ü) dene-di.*
Every actor a costume(-ACC) try-PST.3SG
'Every actor tried a costume.'

(7) *Ali sevgili-sin-e bir yüzüğ(-ü) ver-mek iste-di.*
Ali girlfriend-POSS.3SG-DAT a ring(-ACC) give-INF want-PST.3SG
'Ali wanted to give a ring to his girlfriend.'

In (5), the case-marked direct object *bir sandalyeyi* ('a chair') signals that the speaker has a referential intention with respect to the referent and can identify this referent. In (6), the case marked *bir kostümü* ('a costume') has wide scope with respect to the universal quantifier phrase *her oyuncu* ('every player'). In (7), the case-marked indefinite *bir yüzüğü* ('a ring') has a referential reading with respect to the intensional verb 'want'. While there is quite a lot of research on the sentence behavior of DOM in Turkish (von Heusinger and Kornfilt 2005, Kelepir 2001, Özge 2011 a.o.), there are only very few studies on the discourse behavior of DOM.

2.1 The corpus and the annotation pipeline

In our study we used 21M word *Milliyet Corpus* (Hakkani-Tür, Oflazer, and Tür 2000), which consists of news articles from the daily *Milliyet*, collected from the web in the late 90s. We morphologically analyzed and disambiguated the corpus using the Bosphorus University Morphological Parser and Disambiguator (Sak, Güngör, and Saraçlar 2008). We then syntactically parsed the corpus using the dependency parser Maltparser (Nivre 2008) using the parsing model

of Eryiğit, Nivre, and Oflazer (2008) for Turkish to obtain a treebank. We wrote a workbench in Python that searches the treebank for discourses including a sentence where there is an indefinite noun phrase at the pre-verbal position. The workbench also had a user interface for filtering tasks that required human intervention.

As the annotation workbench we used MMAX2 (Müller and Strube 2006), which is an XML-based tool for specifying annotation schemes, tokenizing input files and serving as a graphical user interface that enables annotators to visually code co-reference relations. We wrote another program that takes MMAX2 output and computes chance-corrected inter-annotator agreement (Artstein and Poesio 2008) and tabulates results for further analyses. We trained two human annotators who manually annotated the automatically selected discourses. We used the 7M portion of the corpus for annotator training. The results we report below are drawn from the remaining 14M part. All the annotated data and tools are open to access at http://users.metu.edu.tr/umozge/var/annotation.tar.gz; browse to 'Phase2' folder for the data reported in this chapter.

3 Backward linking

3.1 Predictions

It has been generally thought that DOM on indefinite direct objects is an indication of some type of linking to preceding discourse (Nilsson 1985). It was Enç (1991) who first articulated a formal proposal concerning the type of linking the DOM indefinites in Turkish have with respect to the preceding discourse. Her account is best explained over an example. In (8) we have a discourse opener followed by one of the minimal pairs given in (8a,b). The pairs differ in the presence versus absence of the accusative case morpheme suffixed to the indefinite object *iki kız* ('two girl').

(8) *Odam-a* *birkaç* *çocuk* *girdi.*
my room-DAT few child entered
'A few children entered my room.' (Enç 1991: 6, ex. 16)

a. *İki* *kız-ı* *tanıyordum.*
two girl-ACC knew.1SG
'I knew two girls.' (Enç 1991: 6, ex. 17)

b. *İki kız tanıyordum.*
two girl knew.1SG
'I knew two girls.' (Enç 1991: 6, ex. 18)

Enç crucially observes that it is with, and only with, the accusative marked version (8a) that the two girls introduced in the discourse belong in to the set of children introduced prior to that point.

Let us look in more detail into what Enç (1991) means by discourse-linking. Enç follows the dynamic semantics tradition of Kamp (1981) and Heim (1982) and takes noun phrases to be linguistic devices that introduce indices (or variables) into the discourse model. However, in Enç's formulation, an NP introduces two indices instead of one as in the standard case. One index serves the function of the standard index, namely it stands for the discourse referent contributed by the NP. The extra index Enç added to the standard model stands for the superset that the discourse referent comes from. Enç holds that the standard dichotomy between definiteness (familiarity) and indefiniteness (novelty) holds for the both indices. For instance, if an NP's both indices are definite, it is a proper definite NP; if its both indices are indefinite, it is a proper indefinite NP. An interesting case which is crucial for Enç's proposal is the case where the discourse referent index is indefinite, while the superset index is definite. Enç takes such NPs to be discourse-linked, a notion she relates to Pesetsky (1987). Therefore, the function of the Turkish accusative marker in the domain of indefinite NPs is to reliably indicate that the discourse referent introduced by the indefinite NP, although itself novel, comes from a familiar set of discourse referents established in the previous discourse. Therefore, we can formulate the first hypothesis:

H1 DOM with indefinite direct objects signals that the new discourse referent is linked to an already established discourse referent by a part-whole relationship (or some other inferred relation.)

One point that needs attention is that the discourse function of DOM is available only when the marker is *not* required by an independent morphosyntactic reason (von Heusinger and Kornfilt 2005). For instance, genitive possessive constructions, although they can be indefinite, obligatorily bear the accusative marker, as can be observed in (9).

(9) a. *Ceren Ahmet-in bir fotoğraf-ın-ı ar-ıyor.*
Ceren A.-GEN a photo-POSS.3SG-ACC seek-PROG.3SG
'Ceren is looking for a photo of Ahmet.' (both specific and non-specific interp.)

b. **Ceren Ahmet-in bir fotoğraf-ı ar-ıyor.*
Ceren A.-GEN a photo-POSS.3SG seek-PROG.3SG

Apart from this, Enç's (1991) double implication between the accusative marker and her discourse-linking has been challenged in both directions, on the grounds that there are (i) accusative marked indefinites entirely new to the discourse and (ii) non-case-marked indefinites linked to the previous discourse (see Taylan and Zimmer 1994, Zidani-Eroğlu 1997, Kelepir 2001, von Heusinger and Kornfilt 2005, Kılıçaslan 2006, İşsever 2003, Nakipoğlu 2009, Özge 2011 a.o.). However, all these studies were based on hand-picked examples constructed by the authors. Therefore, there is a need for a systematic testing of Enç's (1991) proposal. The present study aims to respond to this need. In other words, we aim to put Enç's (1991) and her distractors' judgments into a quantitative test.

With the present study we also aim to contribute to the recently flourishing corpus research on Turkish discourse functions and structure (e.g. Aktaş, Bozsahin, and Zeyrek 2010; Zeyrek et al. 2013 on discourse connectives; Acartürk and Çakır 2012 on referring expressions in situated dialogs).

3.2 Search and annotation on backward linking

In order to test Enç's (1991) proposal formulated in H1 that DOM implies discourse-linking in the sense given in the previous section, we searched the corpus for accusative-marked indefinite objects that appear at the immediately pre-verbal position. We filtered the 4981 tokens we retrieved from the dependency treebank in three ways.

Firstly, we had to filter out a proportion of the retrieved tokens due to errors coming from morphological and syntactic parsing. During the filtering process, 3032 tokens were rejected as errors (either not an indefinite, or not pre-verbal, or both).

Secondly, the tokens where the indefinite is commanded by a nominal or an intensional operator are left out. Only the indefinites that are governed by extensional verbs and do not interact with quantifiers are kept for annotation. The reason for this filtering is that for these cases the marker has the function of inducing flexibility in scope taking (Kelepir 2001; Özge 2011). In these cases, the backward-linking effects are typically missing. For instance, as objects of referentially opaque verbs like *seek*, the accusative induces a referential reading without any implication of familiarity or discourse-linking (Kelepir 2001) as in (10). Therefore, we ruled these non-extensional cases out, in order to better assess the backward discourse-linking effect.

(10) a. *Polis bir çocuğ-u ar-ıyor.*
police a child-ACC seek-PROG.3SG
'The police is looking for a child.' (only specific interpretation, no necessary familiarity effect on the child.)

Thirdly, we filtered out cases where the accusative marker is required for purely grammatical or lexical reasons. For instance, genitive possessive indefinites obligatorily take accusative marking, or the verb *andır* ('resemble') obligatorily takes an accusative marked object, regardless of the sentential and discourse semantic context. However, in cases where optionality is not governed by such a robust principle, we left it to the annotators to decide on the optionality of the case marker in the annotation process.

A set of 1792 tokens were filtered as non-extensional, quantificational or grammatically non-optional, leaving 157 tokens (8% of the all the pre-verbal accusative indefinites retrieved) which were found to be in transparent sentences without extensional or intensional operators.

For each of the 157 tokens we ended up after filtering, we retrieved the entire discourse that the indefinite occurs in. We employed a mixed (links and labels) annotation scheme that has two major levels: discourse and nominal. The first takes entire discourses as markables and is mainly for book-keeping of which discourses are completely annotated, left incomplete or discarded. The nominal level is the central level in our annotation, where the properties of DOM indefinites and their linking relations are coded. Apart from investigating the relation of the indefinite to the preceding discourse, we also aimed to control for categories like the richness of the descriptive content of the indefinite, whether it is in information-structural focus, the place of its referent in an animacy scale, and the indefinite's level of subordination. All these categories are suspected to be effective in the linking pattern of the indefinite. In (11), we provide the annotation categories that are relevant for the backward linking properties, together with their range of values and descriptions.

(11) Annotation categories for accusative marked indefinites:

optional	(yes | no).
descriptive content	(none | adjective | more)
animacy	(human | animate | inanimate concrete object | abstract | undecided)
focal	(yes | no) indicates whether the indefinite receives the main sentential emphasis (nuclear accent) or not.
backward linking	is a link type category.
backward link type	(partitive same | partitive diff | bridged)
bridging type	(part-whole | cause-effect | entity-attribute | other)

We have decided to annotate the following categories: **optional** indicates whether the case marking is optional or not. The optionality test consists in judging whether the same sentence would be acceptable without the accusative marker on the indefinite. The discourse context is not taken into consideration in deciding this category. The cases judged in this annotation category are different from the optionality judgement performed in the filtering phase, where the optionality was ruled out by the grammar in a robust way. The feature **descriptive content** indicates the degree of modification of the nominal head of the indefinite. If there is no modification, the category is annotated as 'none'; if there is only adjectival modification, it gets 'adjective'; and if the modification is clausal, like a relative clause for instance, it gets 'more'. There is evidence (Fodor and Sag 1982) that the more a noun phrase is modified, the more probable it is that it is specific, which means for Turkish that it is case-marked. In order to control for this potential confound, we annotated the amount of descriptive content. The feature **animacy** indicates the animacy of the referent of the indefinite. In Turkish, animacy plays an important role for DOM. Human direct objects are more often case marked than inanimate direct objects (Krause and von Heusinger 2019). We wanted to control for this important parameter for DOM. The first three features provided central parameters for DOM in Turkish. The next features were related to the backward function: **backward linking** is a link type category. The annotators were asked to search for an expression in the discourse prior to the indefinite such that the indefinite is linked to (the referent of) this expression. If the annotator finds such an expression, s/he establishes a link between this expression and the indefinite via MMAX2 interface. This was to establish a discourse link and to identify the anchor of the discourse link in order to establish the type of discourse link with the feature **backward link type,** which is a category conditioned by the previous one. If an annotator links the indefinite to a preceding expression, s/he is requested to indicate the type of the linking. The category is coded 'partitive same' if the linking is as in *the girls – a girl*; it gets 'partitive different', if the linking is as in *the children – a girl*; it gets 'bridged', if the linking is another inferred relation as in *the car – a tyre*. The final feature was **bridging type,** which was to annotate the type of inference between the anchor and the indefinite: If the annotator codes the previous category as 'bridged', then s/he is requested to indicate the type of the bridging relation. A 'part-whole' relation holds, e.g., in *book – a page*, where the indefinite is part of the anchor expression. The 'cause-effect' value covers cases where the anchor causes an effect, such as in *the fire – a victim*. The 'entity-attribute' value covers cases like *movie in theater – a screening*, and finally the category of 'other' for cases that do not fit into any of the above.

Table 1 provides a sample annotation for backward linking:

Table 1: Sample text for annotation (backward linking).

TBMM İnsan Hakları Komisyonu'ndaki tartışmaları okuyunca, "Pes doğrusu" demekten kendimi alamadım. Bir İnsan Hakları Komisyonu düşününüz ki, işkencenin tartışılmasından rahatsız olan üyeler var. Neymiş, **işkence olaylarının** Komisyon'da tartışılması turizmimizi olumsuz yönde etkilermiş. CHP Milletvekili **bir işkence olayını** gündeme getiriyor. Önce, ANAP Milletvekili Yılmaz Karakoyunlu … tepki gösteriyor. Bir başka ANAP Milletvekili olan Süha Tanık'ın sözleri daha çarpıcı: "Bu tür konuların gündeme getirilmesi, turizmimizi baltalar". Bu vekillerimizin mantığına göre, işkence yapıldığı zaman turizm baltalanmıyor da, İnsan Hakları Komisyonu'nda konuşulduğunda baltalanıyor.	When I read the discussions of the Human Rights Commission of Turkish Parliament, I could not help saying "That beats all!" Imagine a Human Rights Commission with members uncomfortable with discussing **torture incidents**. Their reason is that this would badly affect tourism. An MP from CHP brings up **a torture incident**. First, ANAP MP Yılmaz Karakoyunlu… reacts to this. The words of another ANAP MP is even more striking: "Bringing such topics to the commission would hinder tourism." According to the logic of these MPs tourism is hindered not when there is torture but when it is discussed in the commission.

In the text in Table 1, the author introduces a set type referent comprised of incidents of torture in the second sentence. In the next sentence, the author picks one of these incidents by an accusative-marked indefinite *bir işkence olayın-ı* ('a torture incident-Acc'). The NP is annotated as 'backward linked', and the 'backward linking type' receives the value 'partitive-same'. The reason for this choice is that the indefinite and the antecedent expression have the same lexical content; the former is a case-marked indefinite NP, while the latter is a plural NP.

3.3 Results of backward linking

The 157 token discourses, generally newspaper articles, see Table 1, were annotated by two independent annotators. Inter-annotator agreements were generally quite high except for the focal/non-focal category.

First we looked at the optionality of the marker. Out of 157 tokens, the annotators agreed on the optionality category in 156 of the cases. They agreed that the marker is not optional in 3 tokens. Therefore, there were 153 accusative marked indefinites where the marker is agreed to be optional.

Table 2: Inter-annotator agreement for 157 tokens of accusative marked indefinites.

Category	σ	π	κ
Animacy	0.885	0.789	0.790
Backward linking	0.942	0.646	0.647
Descriptive content	0.894	0.892	0.892
Focal or not	0.646	0.185	0.188
Forward linking	0.779	0.756	0.756
Level of subordination	1.0	1.0	1.0
Optionality of case	0.987	0.853	0.853

See Artstein and Poesio 2008 for the three chance-corrected agreement metrics we report here.

For the category of 'animacy', our annotators agreed on 137 tokens out of 153 tokens that were agreed to be optionally case-marked. The results for the category animacy are given in Table 3.

Table 3: Results for the category 'animacy'.

Human	Inanimate Concrete	Abstract	Total
12 (9%)	21 (15%)	104 (76%)	137

For the category of 'descriptive content', our annotators agreed on 142 tokens out of the 153 tokens agreed to be optionally case marked. The results are given in Table 4.

Table 4: Results for the category 'descriptive content'.

None	Adjectival	More	Total
36 (25%)	52 (37%)	54 (38%)	142

For the category 'level of subordination', our annotators agreed on all the 153 tokens that were agreed to be optionally case marked. The results are given in Table 5.

Table 5: Results for the category 'level of subordination'.

Matrix level	1-level embedded	2-level embedded	Total
126 (82%)	24 (16%)	3 (2%)	153

As for the category of 'backward-linking', the annotators agreed on 151 of the 157 tokens of accusative marked indefinites. Out of these 151 cases, 3 cases were agreed to be non-optional case marking and 1 had no agreement on optionality. Therefore, we took 147 tokens into consideration for backward linking. Out of these 147 tokens only 6 were agreed to be backward linked.

Although it is impossible to arrive at any generalization with only 6 tokens, we nevertheless provide the further breakdown of these 6 cases into our annotation categories. As for linking type, 3 was 'partitive with same description', 1 was 'bridging with a part-whole relation' and 2 were 'undecided'; as for descriptive content, 4 were 'none', 1 was adjectival and 1 was 'more'; as for animacy, 4 were abstract, 1 was animate and 1 was inanimate concrete object; as for level of subordination, 4 were matrix level, 1 was one-level, 1 was two-level embedded.

3.4 Discussion on backward linking

We had a high inter-annotator agreement for both the optionality and backward-linking categories (see Table 2) and only the 4% (6 out of 147) of the case marked indefinites were backward-linked to an anchor. All others did not have a discourse relation to a discourse referent introduced earlier. Thus, these results clearly contradict hypothesis H1, which is based on the assumption of Enç (1991) and since then an often repeated assumption in the literature. One possible reservation regarding these results could be that the corpus material we analyzed is biased towards abstract and inanimate concrete direct objects. We do not, however, see why H1 should not also apply to abstract and inanimate concrete direct objects.

Among our annotation categories, the inter-annotator agreement on 'focality' was remarkably poor. This might be due to the fact that newspaper text includes long sentences where the position of the nuclear accent is more likely to be governed by metrical and prosodic concerns – i.e. linguistic aspects that are expected to display individual variability, rather than information-structural requirements.

4 Forward linking

4.1 Predictions

As already mentioned above, certain specially marked indefinites (indefinite *this* in English, *pe*-marking in Romanian) carry a noteworthiness effect, such that the

referents they introduce to the discourse model become more likely to be talked about in the ensuing discourse. Although a similar function for Turkish DOM is hinted at by Taylan and Zimmer (1994), to our knowledge, there has not been any systematic investigation of this idea. Also, some authors have associated DOM with topicality or more generally with a topic shift potential. Case marking on a direct object can – if it is not caused by other grammatical restrictions – indicate that the noun phrase will be picked up in the following discourse. Chiriacescu and von Heusinger (2010) have shown this effect for DOM in Romanian, where *pe*-marked direct objects are more often picked up in the subsequent discourse. Dalrymple and Nikolaeva (2011) call this property of DOM indicating a "secondary topic". This function is demonstrated for Mongolian, an Altaic language typologically similar to Turkish. Guntsetseg (2009) suggests that the decisive factor for DOM in Mongolian is the forward linking property of case-marked indefinite direct objects. The questionnaire contained examples where the object is anaphorically cross-referenced in the next clause (e.g. 'John kissed a girl and she slapped him') and examples without such a relationship (e.g. 'John kissed a girl. James didn't come to school today'). The results marginally suggest that indefinites with case trigger more anaphoric links than indefinites without case.

One obvious metric for this forward-linking effect is the number of referential terms that are anaphoric to the indefinite under discussion. However, explicit anaphoric reference is not the only way to allude to or talk about a referent. One can contribute information regarding a discourse referent by "implicit reference" (Prince 1981:235). For instance, see the examples in Prince (1981:235) for the cataphoric or forward linking potential of indefinite *this*, in (13a) explicit reference, or anaphoric reference, and in (13b) implicit reference or introducing a new (general) topic:

(12) a. 'This fellow I work with -I wouldn't call him militant, but he's perhaps a little more forward than I am – he wouldn't respond if you called him boy. He'd promptly tell 'em ... ' (washroom attendant; (Terkel 1974: 156))
b. 'I been on this one case now about eight months. The problem [in this case] is bad management, not theft ... ' (industrial investigator; (Terkel 1974: 208))

We call the latter type of contribution "elaboration". If Turkish DOM has a forward-linking effect, we expect an increased number of anaphora and elaboration regarding the referent of the indefinite when case-marked.

H2 DOM increases the likelihood of the speaker to continue talking about the referent introduced via the indefinite in question. We refer to this as "forward-linking" function.

4.2 Search and annotation on forward linking

We added the category 'forward linking' to our annotation to indicate the potential of an indefinite direct object to establish an anaphoric chain or to introduce a general new topic the subsequent text is about.

(13) Annotation categories for forward linking:
forward linking none | anaphora | elaboration | elaboration and anaphora

We decided to use four values: **none** if there is no further link in the subsequent discourse; **anaphora** if there is a coreferential expression in the subsequent discourse. We did not count the length of such a referential chain, we just evaluated the fact that there is at least one anaphoric link. We used **elaboration** for discourses that would take up some general topic introduced by the indefinite. And finally, we used **elaboration and anaphora** if the writer both uses at least one expression anaphoric to the indefinite and also elaborates on the referent of the indefinite.

Table 6 provides a sample text for forward linking:

Table 6: Sample text for annotation (forward linking).

[...] Bugün Hacıbektaş ilçesi, yedi yüz yıl öteden seslenen bir büyük düşünürün, büyük hümanistin izini süren yüz binlerce kişinin akınına uğruyor. Hacı Bektaşı Veli'yi anmak isteyenler ve onun aydınlık, insan ve doğa sevgisine dayalı, din, ırk ve cinsiyet ayrımlarını reddeden felsefesinin yolunda yürüyenler bu büyük Veli'yi anmak için Hacıbektaş'ta toplanıyorlar. Biz de onların arasında olacağız. ["Alevi-Bektaşi Düşüncesi ve Çağdaşlık" konulu **bir paneli]** yöneteceğiz. Dileğimizi geri çevirmeyerek ***bu panele*** katılmayı kabul eden değerli dostlarla birlikte Bektaşi değerlerinin çağdaşlık ölçülerine uyumunu tartışacağız. *[Paris'ten gelen Profesör Altan Gökalp, Alman parlamentosu üyesi Cem Özdemir ve eski Kültür Bakanımız Fikri Sağlar'ın katılımcı olduğu **bu panel]*** ilginç geçeceğe benziyor.	Today, the province of Hacı Bektaş will be swarmed by thousands tracing a great thinker, a great humanist calling from seven hundred years back. Those who want to commemorate Hacı Bektaş-i Veli and who follow his path based on love of humanity and nature regardless of any racial or gender differences are gathering in Hacı Bektaş [a village]. We will be among them too. We will direct [**a panel titled** "Alevi-Bektaşi thought and modernity"]. Together with friends who kindly accepted to attend ***this panel***, we will discuss how thoughts of Haci Bektas go with modernity. ***[This panel,*** *which will host Professor Altan Gökalp, German parliamentarian Cem Özdemir and former Minister of Culture Fikri Sağlar]* looks as if it will be interesting.

In the sample in Table 6 there is an accusative-marked direct object (note the case suffix in *panel-i* ('panel-Acc')). In the discourse following the indefinite there are two expressions anaphoric to the case-marked indefinite object, therefore the 'forward linking' category gets the value of 'anaphora'.

Table 7 provides a sample text that illustrates our category of 'elaboration'.

Table 7: Sample text for annotation (elaboration).

HOLLANDA'nın kuzeyindeki Friesland bölgesinde yılbaşı kutlama komitesi **ilginç bir hırsızlık olayını** gerçekleştirdi. Komite, Groningen yakınındaki Tjuchem kasabasındaki bir çiftlikte bulunan 17 ton ağırlığında ve 9 metre boyundaki Lenin heykelini, sahibine haber vermeden 1997'nin son günü gizlice Oosterwolde kasabasına getirmeyi başardı. 1998'in ilk günü kasabanın merkezinde kırmızıya boyanmış meydana, özenle getirilip yerleştirilmiş dev Lenin heykelini görenler gözlerine inanamadılar.	In Friesland area located at north Netherlands, the New Year Celebration Committee has performed **an interesting theft event**. The committee succeeded in bringing the 17 tons, 9 meter Lenin statue located in a farm in Tjuchem near Groningen to Oosterwolde village wihtout any notice of the owner on the last day of 1997. On the first day of 1998, those who saw the Lenin statue brought and put at the center of the village, which was painted in all red, could not believe their eyes.

In the text in Table 7, the author introduces an event referent in the opening sentence. In the rest of the text, the author does not use any term anaphoric to this event referent. However, s/he gives further information on the referent, recounting the details of the introduced event.

4.3 Results of forward linking

The annotators agreed in 128 out of 153 tokens of accusative-marked indefinites agreed to be optionally case marked on their response to the category 'forward-linking'. The inter-annotator agreement was at an acceptable level ($\kappa = 0.756$, see Table 2. The distribution of the responses are given in Table 8.

Table 8: Forward linking of accusative indefinites.

No linking	Anaphora	Elaboration	Elab. + Anaph.	Total
46 (36%)	26 (20%)	52 (41%)	4 (3%)	128

Our data shows that the writers continue to write about the referent of an accusative-marked indefinite in 64% of the cases. In order to be able to judge

whether this tendency to continue to write about the referent of the indefinite is an effect of the marker itself, we prepared a class of zero-marked indefinites for comparison.

One complication in having a class of zero-marked tokens for comparison with the accusative marked ones is the fact that zero-marked indefinites are much more common than accusative-marked ones. It is hard to estimate the ratio of the two, because we do not know the number of zero-marked indefinite objects in our corpus, due to the annotator cost of filtering the errors in the automatically retrieved tokens. However, it is easy to get an idea of the proportion of accusative-marked indefinites to zero-marked ones by looking at the situation for a single verb: While there are 6 tokens of an accusative indefinite as the object of verb *göster* ('show'), the number of tokens where this verb takes a zero-marked indefinite as object is 310.

In order to circumvent this difficulty arising from the disproportionate number of accusative-marked and zero-marked indefinites, we picked a subset of the tokens of accusative-marked indefinites. This subset is formed by excluding the verbs which had less than three tokens of an accusative-marked indefinite object. After this filtering, we were left with a set of 51 tokens of accusative-marked indefinites, with the 13 verbs in (14). We then retrieved 60 tokens of zero-marked indefinites occurring as objects of the same 13 verbs. The filtering principles we used for accusative-marked indefinites – no nominal or intensional operators, optionality of the marker – applied here as well. We annotated this set of zero-marked indefinite tokens with respect to forward linking property.

(14)

aktar	'transmit'	*al*	'take'	*anlat*	'explain/tell'
benimse	'adopt'	*gerçekleştir*	'realize'	*göster*	'show'
kabul et	'accept'	*öldür*	'kill'	*oluştur*	'form'
üstlen	'undertake'				
başlat	'start'				
gündeme getir	'mention'				
ortaya çıkar	'reveal'				

The annotators were in agreement in 56 out of 60 cases (κ=0.772). On the other hand, the annotators agreed on their response to 'forward linking' category in 46 out of 51 tokens of accusative indefinites. The distribution of forward linking properties of zero marked indefinites and the 46 token subset of accusative-marked indefinites is given in Table 9, where we included the 'anaphora + elaboration' responses in 'anaphora'.

Table 9: Forward linking properties of zero-marked and accusative indefinites.

	Zero		Accusative	
No linking	16	29%	16	35%
Anaphora	12	21%	10	22%
Elaboration	28	50%	20	43%
Total	56	100%	46	100%

4.4 Discussion on forward linking

We created two corpora of similar size and structure – one with instances of indefinite direct objects without accusative case and one with instance with accusative case. In all instances, the indefinite was preceded by the indefinite article *bir*. Comparing these two samples suggests that DOM does not induce a forward linking effect on the indefinites it is attached to. At the same time, we observed about 65 to 70% of forward linking for both samples. We think this is an impressively high number and we therefore speculate that these types of examples might not be able to reflect a difference in the forward linking potential, due to a high baseline likelihood of further reference.

5 Conclusion

Differential case marking is a ubiquitous phenomenon in various languages. It is determined by syntactic position, referentiality of the noun, verbal semantics and information structure. The goal of this paper was to investigate whether case marking not only signals a particular sentence semantic behavior in terms of specificity, but also a discourse semantic behavior in terms of discourse prominence. We formulated two hypotheses with respect to the discourse semantic behavior: H1 was based on the assumption that DOM indefinites are backward-linked to an already introduced anchor expression. This hypothesis was initiated by the seminal work of Enç (1991) and supported by subsequent observations. Our second hypothesis concerned the forward linking properties, i.e. the potential to shift a topic and the potential to be the antecedent of an extended anaphoric chain (Taylan and Zimmer 1994). This is supported by observations of DOM in other languages (for Romanian: Chiriacescu and von Heusinger 2010, for Mongolian: Guntsetseg 2009, for a general perspective Dalrymple and Nikolaeva 2011).

In order to test these two hypotheses, we searched and annotated DOM-marked indefinite direct objects with respect to their backward and forward discourse functions in a 14M words newspaper text. Our search and filtering pipeline delivered 157 tokens of accusative-marked indefinite direct object in an extensional context (no nominal or intensional operator commanding the indefinite). In the backward direction, out of 147 tokens that our annotators agreed both on the optionality of DOM-marking and backward-linking type of, only 6 were judged related to an antecedent in the preceding discourse. In the forward direction, in order to establish a basis for comparison, we retrieved and annotated a comparable amount of non-DOM-marked tokens. We did not observe any difference between DOM-marked versus non-DOM marked indefinites in their discourse behavior in the forward direction. Overall, our corpus study contributes negative evidence regarding both the backward and forward linking discourse functions for Turkish DOM. These negative results are supported by a production experiment of Özge, Özge, and von Heusinger (2016), where we asked informants to continue sentences with indefinite direct objects with and without case marking. We then counted the number of anaphoric expressions. There was no significant difference between antecedents with case and antecedents without case. It seems that DOM in Turkish does not contribute to the discourse prominence of the associated referent. This is contrary to DOM in other languages such as Romanian.

The difference between DOM in Turkish and Romanian might be due to the fact that in Turkish DOM is expressed by a case suffix, while in Romanian it is expressed by a free lexeme *pe*. Second, von Heusinger and Bamyacı (2017) have provided evidence that DOM in Turkish is correlated with scopal and referential specificity, but not with epistemic specificity (see von Heusinger 2011, 2019 for the different types of specificity). We speculate that it might be the discourse pragmatic epistemic specificity that influences the discourse prominence of the referent, but not its scopal properties.

Another note is that due to the particular register of newspaper text we have extracted a high proportion of abstract direct objects – a type of direct object that is rarely investigated in the research on DOM in particular and on differential case marking in general. It is yet to be seen whether abstract objects behave similarly to concrete ones regarding the discourse effects of DOM in other languages.

Finally, we assume that the discourse functions of indefinite direct objects are licensed by the (not very frequent) use of the indefinite article *bir*, rather than the case marker. Thus, our corpus search and analyses contribute to the relation of sentence semantics and discourse semantics.

Acknowledgements: We would like to thank two anonymous reviewers for very helpful and constructive comments. We also thank our annotators Gökben Konuk

and Serkan Yüksel for their accurate annotation. We thank Stefanie Dipper, Hans Kamp, Arndt Riester, Duygu Özge, Heike Zinsmeister and an anonymous reviewer for their valuable comments. The research for this paper has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – "Indefinites in Discourse" (HE 6893/14-1) and Projektnummer 281511265 – SFB "Prominence in Language" in the project B04 "Interaction of nominal and verbal features for Differential Object Marking" at the University of Cologne.

References

Acartürk, Cengiz & Perit M. Çakır. 2012. Towards building a corpus of Turkish referring expressions. In Şeniz Demir, İlknur Durgar El-Kahlout & Mehmet Uğur Doğn (eds.), *Proceedings of the 1st Workshop on Language Resources and Technologies for Turkic Languages*, 1–5.

Aktaş, Berfin, Cem Bozsahin & Deniz Zeyrek. 2010. Discourse relation configurations in Turkish and an annotation environment. In *Proceedings of the 4th Linguistic Annotation Workshop* 202–206. Stroudsburg, PA: Association for Computational Linguistics.

Artstein, Ron & Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics* 34 (4). 555–596.

Chiriacescu, Sofiana & Klaus von Heusinger. 2010. Discourse prominence and *pe*-marking in Romanian. *International Review of Pragmatics* 2. 298–332.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

Deichsel, Annika & Klaus von Heusinger. 2011. The cataphoric potential of indefinites in German. In Iris Hendrickx, Sobha Lalitha Devi, António Branco & Ruslan Mitkov (eds.), *Anaphora processing and applications, DAARC* 2011, *Lecture Notes in Computer Science*, 144–156. Heidelberg: Springer.

Enç, Mürvet. 1991. The Semantics of specificity. *Linguistic Inquiry* 22. 1–25.

Eryiğit, Gülşen, Joakim Nivre & Kemal Oflazer. 2008. Dependency parsing of Turkish. *Computational Linguistics* 34 (3). 357–389.

Fodor, Janet & Ivan Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5. 355–398.

Guntsetseg, Dolgor. 2009. Differential object marking in (Khalkha-)Mongolian. In Ryosuke Shibagaki & Reiko Vermeulen (eds.), *Proceedings of the 5th Workshop on Formal Altaic Linguistics (WAFL 5)*, 115–129. Cambridge, MA: MIT Working Papers in Linguistics.

Hakkani-Tür, Dilek Z., Kemal Oflazer & Gökhan Tür. 2000. Statistical morphological disambiguation for agglutinative languages. In *Proceedings of the 18th Conference on Computational Linguistics – Volume 1, COLING* '00, 285–291. Stroudsburg, PA: Association for Computational Linguistics.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases in English*. Amherst, MA: University of Massachussetts, Amherst dissertation.

von Heusinger, Klaus. 2011. Specificity. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics. An international handbook of natural language*

meaning, vol. 2 (HSK 33.2), 1025–1058. Berlin: de Gruyter. doi:https://doi.org/10.1515/9783110255072.1025.

von Heusinger, Klaus. 2019. Indefinites and specificity. In Jeanette Gundel & Barbara Abbott (eds.), *Oxford handbook of reference*, 146–167. Oxford: Oxford University Press

von Heusinger, Klaus & Elif Bamyacı. 2017. Specificity effects of Turkish differential object marking. In Leyla Zidani-Eroğlu, Matthew Ciscel & Elena Koulidobrova (eds.), *Proceedings of the 12th Workshop on Altaic Formal Linguistics (WAFL12)*, 309–319. Cambridge, MA: MIT Working Papers in Linguistics.

von Heusinger, Klaus & Umut Özge. submitted. Inferrable and partitive indefinites in topic position. In Anke Holler, Katja Suckow, Barbara Hemforth & Israel de la Fuente (eds.), *Information structuring in discourse*. Leiden: Brill.

von Heusinger, Klaus & Jaklin Kornfilt. 2005. The case of the direct object in Turkish: Semantics, syntax and morphology. *Turkic Languages* 9. 3–44.

Ionin, Tania. 2006. *This* is definitely specific: specificity and definiteness in article systems. *Natural Language Semantics* 14. 175–234.

İşsever, Selçuk. 2003. Information structure in Turkish: the word order-prosody interface. *Lingua* 113 (11). 1025–1053.

Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen, and Martin Stokhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Mathematical Center.

Kamp, Hans. 2014. *Dividing the province of indefinite noun phrase uses into three parts*. Ms. Universität Stuttgart / University of Texas at Austin.

Kamp, Hans & Agnes Bende-Farkas. 2018. Epistemic specificity from a communication-theoretic perspective. *Journal of Semantics* 36 (1). 1–51. doi:https://doi.org/10.1093/jos/ffy005.

Kelepir, Meltem. 2001. *Topics in Turkish syntax: Clausal structure and scope*. Cambridge, MA: MIT dissertation.

Kılıçaslan, Yılmaz. 2006. A situation-theoretic approach to case marking semantics in Turkish. *Lingua* 116. 112–144.

Krause, Elif & Klaus von Heusinger. 2019. Gradient effects of animacy on differential object marking in Turkish. *Open Linguistics* 5. 171–190.

MacLaran, Rose. 1982. *The Semantics and pragmatics of the English demonstratives*. Ithaca, NY: Cornell University dissertation.

Müller, Christoph & Michael Strube. 2006. Multi-level annotation of linguistic data with MMAX2. In Sabine Braun, Kurt Kohn & Joybrato Mukherjee (eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods*, 197–214. Frankfurt a.M.: Peter Lang.

Nakipoğlu, Mine. 2009. The semantics of the Turkish accusative marked definites and the relation between prosodic structure and information structure. *Lingua* 119. 1253–1280.

Nilsson, Birgit. 1985. *Case marking semantics in Turkish*. Stockholm: University of Stockholm dissertation.

Nivre, Joakim. 2008. Algorithms for deterministic incremental dependency parsing. *Computational Linguistics* 34 (4). 513–553.

Özge, Umut. 2011. Turkish indefinites and accusative marking. In Andrew Simpson (ed.), *Proceedings of the 7th Workshop on Altaic Formal Linguistics*, 253–267. Cambridge, MA: MIT Working Papers in Linguistics.

Özge, Umut, Duygu Özge & Klaus von Heusinger. 2016. Strong indefinites in Turkish, referential persistence, and salience structure. In Anke Holler & Kaja Suckow (eds.), *Empirical*

perspectives on anaphora resolution, 169–191. Berlin: de Gruyter. doi:https://doi.org/10.1515/9783110464108-009.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland & Alice ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, MA: MIT Press.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.

Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In William C. Mann & Sandra A. Thompson (eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text*, 295–325. Amsterdam: John Benjamins.

Sak, Haşim, Tunga Güngör & Murat Saraçlar. 2008. Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In Berg Nordstöm & Aarne Ranta (eds.), *Advances in natural language processing*, 417–427. Berlin, Heidelberg: Springer.

Seidel, Elyesa. 2019. *Pseudo-incorporation and event structure*. Cologne: University of Cologne dissertation.

Taylan, Eser & Karl Zimmer. 1994. Case marking in Turkish indefinite object constructions. In Kevin E. Moore, David A. Peterson & Comfort Wentum (eds.), *Proceedings of the 20th Annual Meeting of Berkeley Linguistics Society*, 547–553. Berkeley, CA: Berkeley Linguistic Society.

Terkel, Studs. 1974. *Work: Working people talk about what they do all day and how they feel about what they do*. New York: Random House.

Zeyrek, Deniz, Işın Demirşahin, Ayışığı B. Sevdik-Çallı & Ruket Çakıcı. 2013. Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. *Discourse and Dialogue* 4(3). 174–184.

Zidani-Eroğlu, Leyla. 1997. *Indefinite noun phrases in Turkish*. Madison, WI: University of Wisconsin-Madison dissertation.