Differential Object Marking in Turkic and Persian as a Contact Phenomenon
Author(s): Greg Key
General Session and Thematic Session on Language Contact
Editors: Kayla Carpenter, Oana David, Florian Lionnet, Christine Sheil, Tammy Stark, Vivian Wauters

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via [http://linguistics.berkeley.edu/bls/](http://linguistics.berkeley.edu/bls/).

*The Annual Proceedings of the Berkeley Linguistics Society* is published online via eLanguage, the Linguistic Society of America's digital publishing platform.
Differential Object Marking in Turkic and Persian as a Contact Phenomenon

GREG KEY

The University of Arizona

1. Object Marking in Turkish, Persian, and the Languages of Iranian Azerbaijan.¹

Despite the absence of genetic affiliation, Turkish (SW Turkic) and Persian (SW Iranian) have nearly identical differential object marking patterns. Herein it is proposed that this is due in part to contact between Persian and Azerbaijanian, which is closely related to Turkish, and that Iranian Azerbaijan is an isogloss for this feature. The tableau of evidence is a large puzzle only a few pieces of which have been filled in. In this work, I present several of those pieces. Part 1 contains synchronic evidence of object marking patterns in various languages both inside and outside the proposed isogloss, while Part 2 contains the result of a study comparing object marking in Old Anatolian Turkish and Classical Persian manuscripts.

1.1 The Turko-Persian Pattern of Differential Object Marking (DOM)

Differential Object Marking (DOM) is the morphological marking of some direct objects and not of others, based on hierarchies such as animacy and referentiality (Bossong 1985, 1981; Aissen 2003). The Persian marker is –râ in the formal language, colloquially realized as –ro following a vowel and –o following a consonant. The Turkish object marker is –I following a consonant and –yI following a vowel, where I represents a high vowel realized as front or back, rounded or unrounded, according to the rules of Turkish vowel harmony.

The following marking pattern holds in both languages, regardless of the animacy of the object. Definite objects are obligatorily marked. Indefinite objects

¹ Thanks to Davoud Zamani for Iranian Azerbaijanian, and to Simin Karimi for Persian grammaticality judgments.
are usually unmarked, but may be marked under certain conditions.\textsuperscript{2} Bare nominal objects (having no determiner or indefiniteness morphology) have a definite singular interpretation when marked, and a \textit{kind-level} interpretation when unmarked. A \textit{kind-level} noun (Karimi 2005) is unspecified for number, and non-referential (also called \textit{categorial} (Göksel & Kerslake 2005)).

Feature 1: Definite objects are obligatorily marked.\textsuperscript{3}

\textbf{Persian}

(1) \textit{Kimea ket\text{"a}b-o} xund.
\textit{Kimea book-OM} read
‘Kimea read the book.’

\textbf{Turkish}

(2) \textit{Ay\text{"a}se kitab-t} okudu.
\textit{Ay\text{"a}se book-OM} read
‘Ay\text{"a}se read the book.’

Feature 2: Indefinite objects are usually unmarked, but may be marked.

\textbf{Persian}

(3) a. \textit{Kimea ye d\text{"a}st\text{"a}n-(i)-_ goft.}
\textit{Kimea one story-IND-Ø} said
‘Kimea told a story.’

b. \textit{Kimea ye d\text{"a}st\text{"a}n-i-ro goft ke az to \text{"s}enide bud.}
\textit{Kimea one story-REL-OM told that from you heard was}
‘Kimea told a story that (she) had heard from you.’ (Karimi 2005: 27)

\textbf{Turkish}

(4) a. \textit{Bazen masa-ya bir \text{"o}rt\text{"u}- yay-ar-di-k.}
\textit{sometimes table-DAT one cloth-Ø spread-AOR-PST-2PL}
‘Sometimes we would spread a cloth on the table.’

\textsuperscript{2} Marking of indefinites is claimed to be based on the specificity of the object (Enç 1991 for Turkish, Karimi 2005 for Persian). However, not all authors are in agreement on this, nor indeed on the proper characterization of ‘specificity’ (see Ghomeshi 1997, von Heusinger & Kornfilt 2005). Given the difficulty of characterizing specificity independently in either language, I will not address it in the present cross-linguistic study.

\textsuperscript{3} Abbreviations: ABL ablative, AOR aorist, DAT dative, DUR durative, ERG ergative, GEN genitive, IMP imperative, IND indefinite, NOM nominative, OM object marker, POSS possessive, DEF definite, NEG negative, OBL oblique, OPT optative, PERF perfect, PL plural, PRET preterit, PST past, SG singular.
Differential Object Marking in Turkic and Persian as a Contact Phenomenon

b. Bazen masa-ya Ayşe-nin bize Meksika-dan
sometimes table-DAT Ayşe-GEN we-DAT Mexico-ABL
getir-diğ-i bir örtü-yü yay-ar-di-k.
bring-PRT-3SG one cloth-OM spread-AOR-PST-2PL
‘Sometimes we would spread on the table a cloth that Ayşe had
brought us from Mexico.’ (Göksel and Kerslake 2005: 375-376)

Feature 3: Bare nominal objects have a definite singular interpretation when
marked, and a kind-level interpretation (unspecified for number, and non-
referential) when unmarked. For convenience’s sake, the kind-level object
ketâb/kitap is translated as ‘a book/books’; however, it should be noted that ‘book’
as a kind-level object does not imply that any book was read in its entirety, and
hence predicates taking kind-level objects are atelic.

Persian
Kimea book-OM read
‘Kimea read the book.’
b. Kimea ketâb-ø xund.
Kimea book-ø read
‘Kimea read a book/books.’

Turkish
(6) a. Ayşe kitab-ı okudu.
Ayşe book-OM read
‘Ayşe read the book.’
b. Ayşe kitap-ø okudu.
Ayşe book-ø read
‘Ayşe read a book/books.’

1.2 DOM in Other Turkic and Persian Languages

Is this shared pattern due to contact? Although Turkey and Iran are
contiguous, Turkish and Persian do not share a contact area. Literary Turkish of
the Ottoman Empire was under influence from literary Persian, but the reverse
was not true. However, Persian is in contact with Azerbaijani (SW Turkic), a
close relative of Turkish. The contact zone is Iranian Azerbaijan (in northwestern
Iran). The Azerbaijani language spoken in Iran has the same DOM pattern.
Feature 1: Definite objects are obligatorily marked.

(7) Aisha kitab-ı okhudu.
Aisha book-OM read
‘Aisha read the book.’

Feature 2: Indefinite objects are usually unmarked, but may be marked.

(8) Iran barasında bir kitab-/kitab-ı akhtar-ı r-am.
Iran about a book-Ø/book-OM search-PRES-1SG
‘I’m looking for a book about Iran.’

Feature 3. Bare nominal objects have a definite singular interpretation when marked, and a kind-level interpretation when unmarked.

(9) a. Aisha kitab-ı okhudu.
Aisha book-OM read
‘Aisha read the book.’

b. Aisha kitab-ı okhudu.
Aisha book-Ø read
‘Aisha read a book/books.’

Since the contact area for Persian and Azerbaijani is Iranian Azerbaijan, I propose that this area is an isogloss for the DOM pattern identified herein. As a first step in investigating this proposal, it is necessary to determine the distribution of this pattern both inside and outside of the proposed isogloss. The present study represents the rudimentary beginnings of such a project.

The languages spoken in Iranian Azerbaijan include Azerbaijani (SW Turkic), Eastern Armenian (precise affiliation within the Indo-European family unclear), and the Iranian languages Persian (SW), Tatic (Tati, Taleshi), Gilaki, and Mazanderani (NW).

Eastern Armenian shows evidence of the pattern. In the dialect spoken in Iran, the morpheme ə (/n/ after vowels) differentially marks direct objects (Megerdoomian 2008). In traditional grammars, this is identified as the definite article. According to Megerdoomian, it is the marker for inanimate objects. The form –an is used for animates but apparently has the same differential properties with regard to definiteness. (The glosses have been slightly modified; in particular, Megerdoomian’s gloss ACC (accusative) has been changed to OM.) ‘YES’ indicates a feature’s presence, while ‘NO’ indicates its absence.
Feature 1: YES

(10) Ara-n girk'-ə ayr-ets
    Ara-NOM book-OM burned
    ‘Ara burned the book.’ (Megerdoomian 2008)

Feature 2: YES

(11) Ara-n mi girk'-_/girk'-ə ayr-ets
    Ara-NOM one book-ø/book-OM burned

Feature 3: YES

(12) a. Sirun-ə xandzor-ə ker-av
    Sirun-NOM apple-OM eat-AOR/3SG
    ‘Sirun ate the apple.’

b. Sirun-ə xandzor-_ ker-av
    Sirun-NOM apple-OM eat-AOR/3SG
    ‘Sirun ate an apple/apples.’ (Megerdoomian 2008)

At present I have scant data on DOM in NW Iranian languages, limited to examples in Bossong (1985) and Windfuhr (2009), which for the most part provide information on Feature 1 only. I will therefore limit the discussion of these languages to this feature.

Since inanimates are lower on the animacy hierarchy than animates, the marking of definite inanimate objects entails the marking of definite animates (Aissen 2003). Therefore, examples of marked inanimate objects are taken as evidence that marking is possible for all definite objects, regardless of animacy. Furthermore, the absence of any examples with unmarked definite objects leaves open the possibility that the marking of definite objects is obligatory. Such languages are tentatively considered to have Feature 1, the possibility that they exhibit full pattern remaining open (including Features 2 and 3).

Examples with unmarked definite objects are taken to mean that the language in question does not exhibit Feature 1, and hence does not exhibit the Turko-Persian DOM pattern. For Tati, Taleshi, Gilaki and Mazanderani, Bossong gives examples of unmarked indefinite objects, both animate and inanimate, showing that these languages have DOM. He provides no examples with unmarked definite direct objects, leaving open the possibility that these languages have Feature 1. (Again, the glosses have been modified slightly for consistency.)

Feature 1: Definite objects are obligatorily marked.

Tati

(13) ä täxtä musmar-ää män bā-kālbāti-raz vākānd-yīm
    from board nail-OM I to.pliers.with pull.out-1SG
    ‘I pulled the nail out of the board with a pair of pliers.’ (Bossong 1985: 56)
There are no definite objects among Bossong’s examples of unmarked objects.

Taleshi
(14) țăbut-e bo-nia zamin.
coffin-OM CONJ-sets ground
‘…sets the coffin on the ground.’

Gilaki
(15) səməvər-a ātəš bu-kun
Samovar-ACC fire CONJ-do
‘Light the samovar!’

Mazanderani
(16) jək zan in harf-râ bə-šni-a
one woman this word-OM PRET-hear-3SG
‘A woman heard these words.’

Kurdish varieties, on the other hand, which are not spoken in Iranian Azerbaijan, lack the pattern. Sorani Kurdish entirely lacks morphological marking of objects altogether (Thackston 2006b). Kurmanji Kurkish, on the other hand, exhibits split ergativity. In the present tense, feminine nouns are marked OBL regardless of definiteness or specificity (i.e., marking is non-differential). Masculine nouns are not marked unless preceded by a demonstrative (Thackston 2006a).

1SG NEG-go-1SG doctor 1SG medicine-Ø PRV-NEG-drink-1SG
‘I’m not going to the doctor. I’m not taking medicine. I don’t want stitches, either.’ (Thackston 2006a: 35)

Vafsi (NW Iranian), spoken outside of Iranian Azerbaijan, has DOM but not the Turko-Persian variety. It is also split ergative, and has DOM in the present tense only, where animate, specific objects are marked oblique (Stilo 2004).
Feature 1—NO

Definite animate: marked
(19)  tæ  in  xær-i  næ-ruš-i?
you this donkey-OM  NEG-SELL-2SG
‘Won’t you sell this donkey?’

Definite inanimate: not marked
(20)  gázæ-õ  ú-gur
pincers-ø  PVB-take
‘Get the pincers.’ (Stilo 2004: 243)

The NW Iranian languages that appear to exhibit Feature 1 have a contiguous
distribution in northwest Iran (see Figure 2). NW Iranian languages that clearly
lack Feature 1 are found along this region’s southern and western periphery. Note
that the languages lacking this feature are surrounded by languages that have it,
not only in NW Iran, but also Persian to the south and Turkish to the west. This
distribution is consistent with a model wherein NW Iranian initially lacked the
feature, but NW Iranian languages in Iranian Azerbaijan and the Caspian region
later acquired it. It is also noteworthy that the these languages pattern together in
seven of the eight isoglosses discussed in Stilo (2005).

It is also necessary to determine the DOM patterns in Turkic and Iranian
languages farther afield from Iranian Azerbaijan. Here especially, an immense
amount of work remains to be done.

Uzbek (SE Turkic) appears to have Features 1 and 2 (data from Raun 1969). I
do not currently have information on Feature 3.

Feature 1
(21)  kitp-ni  oqiýdi
book-OM  reads
‘He reads the book.’

Feature 2
(22)  men-ga  bir  stakan suw  ber-iñ!
1SG-DAT a  glass water  give-IMP
‘Give me a glass of water!’ (Raun 1969: 20)

Raun points out that ‘-ni may also be used to denote a less definite, or even
indefinite object’ (p. 20), although he gives no such examples. This indicates that
Uzbek has Feature 2 as well.

As for geographically removed Iranian languages, Tajik, a variety of Persian
spoken in Tajikistan and Uzbekistan also has Features 1 and 2 (Windfuhr & Perry
2009: 485). It is difficult to know exactly what to make of this. On the one hand,
Tajik is a variety of Persian, and hence may inherit this feature from Persian
generally. On the other hand, there is a plethora of contact situations in Central Asia, and there may be more than one isogloss for this trait.

Feature 1
(23) kitob-ro xarid-am
book-OM bought-1SG
‘I bought the book.’

Feature 2
(24) (yak) zan[-e]-ro did-am
one woman-IND-OM saw-1SG
‘I saw a certain woman.’

On the other hand, Hindi/Urdu (Indo-Iranian, Indic branch; India, Pakistan) lacks Feature 1.

(25) a. Ravi-ne kacca a kelaa-_ kaaṭaa.
Ravi-ERG unripe banana-Ø cut.PRF
‘Ravi cut the unripe banana.’

b. Ravi-ne kacce kele-ko kaaṭaa.
Ravi-ERG unripe banana-OM cut.PRF
‘Ravi cut the unripe banana.’ (Mohanan 1994a: 87-88)

2 Textual Evidence from Early New Persian and Old Anatolian Turkish

DOM patterns in the historical predecessors of Turkish/Azerbaijanian and Persian are another important piece of the puzzle. In Old Turkic (8th-13th centuries AD), non-specific objects are never marked, but marking for definites is optional (Erdal 2004). In Old Persian (6th-4th centuries BC), the accusative marker was not differential (i.e., it was used on all direct objects). Middle Persian (300 BC-AD 900) marked some objects, but it is not clear if its use was differential (Brunner 1977).

In the following, a study of two medieval texts, one Early New Persian (ENP) and one Old Anatolian Turkish (OAT), is presented. These were analyzed to determine the distribution of object markers on direct objects based on objects’ definiteness and animacy. ENP data were collected from a facsimile of an early 14th-century copy of the Nasihat-nāme of Onsorolma‘āli Kaykâ’us (Leiden Codex, dated AH 719/AD 1319), while OAT data were collected from a late-14th-century copy of a translation of the Nasihat-nāme (Birnbaum MS T 12, date missing). The original Persian was composed in the late 11th century. The Turkish manuscript is considerably shorter due to lacunae in the extant manuscript, and consequently fewer data were collected from it. Only the narrative portions of the texts were included.
In some sense, the texts are less than ideal for comparison, since they are from different centuries. However, the fact that the OAT text is a translation of the ENP text has some advantages. First, it allows us to compare identical examples in the two languages. Next, it makes clear that, as seen below, the Turko-Persian pattern found in OAT is not a translation effect.

Table 1. The distribution of marked and unmarked objects in the PQN

<table>
<thead>
<tr>
<th></th>
<th>+human</th>
<th>−human/ +animate</th>
<th>−animate</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>=rā</td>
<td>ø</td>
<td>=rā</td>
<td>ø</td>
</tr>
<tr>
<td>Pronoun</td>
<td>54</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Proper Name</td>
<td>18</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Definite Common NP</td>
<td>23</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Indefinite Common NP</td>
<td>9</td>
<td>23</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>104</td>
<td>25</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Combined Total</td>
<td>129</td>
<td>12</td>
<td>373</td>
<td>141</td>
</tr>
<tr>
<td>Doubtful definiteness</td>
<td>3</td>
<td>1</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. The distribution of marked and unmarked objects in the TQN

<table>
<thead>
<tr>
<th></th>
<th>+human</th>
<th>−human/ +animate</th>
<th>−animate</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>=(y)I</td>
<td>ø</td>
<td>=(y)I</td>
<td>ø</td>
</tr>
<tr>
<td>Pronoun</td>
<td>45</td>
<td>0</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>Proper Name</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Definite Common NP</td>
<td>14</td>
<td>5</td>
<td>105</td>
<td>7</td>
</tr>
<tr>
<td>Indefinite Common NP</td>
<td>5</td>
<td>16</td>
<td>22</td>
<td>116</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>16</td>
<td>152</td>
<td>123</td>
</tr>
<tr>
<td>Combined Total</td>
<td>89</td>
<td>8</td>
<td>275</td>
<td>372</td>
</tr>
<tr>
<td>Doubtful definiteness</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Human definite nouns were as a rule marked in both texts. There are two exceptions in the PQN (see Key 2008 for discussion) and none in the TQN. In both, human indefinites were sometimes marked but more frequently unmarked. The main divergence between the texts is in the marking of inanimate definite common nouns, which, as mentioned in Section 1, is an important diagnostic for Feature 1.

In the PQN, only 9 out of 186 inanimate definite objects are marked, while in the TQN 105 out of 112 inanimate definite objects are marked.
This shows that the EPN variety exemplified in the PQN clearly lacked Feature 1, and hence the Turko-Persian pattern as a whole. In the TQN, on the other hand, the fact that 7 out of 112 inanimate definites (6.25%) were unmarked shows that the pattern was not absolute, but that marking was overwhelmingly the rule.

With respect to Feature 2, the PQN treated animate indefinites much as modern Persian does, but not inanimate indefinites, of which only 4 out of 157 were marked (2.5%). The TQN, on the other hand, marked such objects at a much higher rate, 22 out of 116 (19%).

Animate indefinites:

PQN

(28) mo’tasim vaqti mojrem-i-râ piš-e xwiš kardan hamî mutasim when criminal-IND-OM front-EZ self make DUR farmud commanded

‘When Mutasim commanded that [they] bring a criminal before him...’

(PQN 31b)

TQN

(29) mu’tasim...bir sučlu-yî buyur-dî kim getür-e-ler mutasim... a criminal-OM command-PST that bring-OPT-PL

‘Mutasim...commanded that they bring a criminal forward’ (TQN 48a)
Inanimate indefinites:

PQN

(30) šenid-am ke vaqti sâheb esmâʾilben ʿubbâd nân-_ hamī
declared-1SG that when sir Ismail bin ʿubbad bread-Ø CNT
xord bâ nadimān-e xwiš mard-i loqma-i-
ate with companions-Ø own morsel-Ø
az kāse bar-dāšt from bowl picked.up
‘I heard that when sir Ismail Ubbad was eating bread with his
companions, a man picked up a morsel from the bowl’ (PQN 27a)

TQN

(31) bir gün ismâʾil ʿubbâd yoldaš-lar-î-le yiyesi
one day Ismail Ubbad companion-PL-3SG-with food
ye-r-idi bir kiši bir loqma-yî götür-icek daxī
eat-AOR-PST one person a morsel-OM pick.up-when also
bir qīl bile yapīš-dī.
a hair with stick-PST
‘One day Ismail Ubbad was eating with his companions. And when
one person picked up a morsel, a hair stuck [to it].’

As for Feature 3, there is a dearth of evidence of kind-level objects in both
manuscripts. There is a singled shared example.

PQN

(32) xatt-e mozavvar-_ kard-i
letter-Ø forged-Ø did-HAB
‘He wrote forged letters.’ (PQ 100a)

TQN

(33) sâhib bun-i iṣit-di kim bu tezvîr biti-
master this-OM hear-PST that this forgery letter-Ø
yaz-ar devû
write-AOR saying
‘The master heard this, that he wrote forged letters.’ (TQ 119a)

Extreme caution must be exercised in interpreting these results. Although the
EPN manuscript lacks the Turko-Persian DOM pattern, while the OAT
manuscript for the most part exhibits it, the former is three centuries older than
the latter, and so this cannot on its own be taken to mean that Turkic had the
pattern before Persian. However, the following observations can be made. First,
the pattern found in the PQN, where definites are virtually always marked and
indefinites sometimes marked only if they are animate is very similar to patterns
found in genetically related languages such as Vafsi and Hindi/Urdu. Next, OAT
clearly exhibited something very close to the modern Turko-Persian pattern, and this was not a translation effect, since the ENP of the source did not have the feature.

3 Summary and Further Proposal

The three features that make up the DOM pattern found in Turkish (SW Turkic) and Persian (SW Iranian) are also found in Azerbaijanian (SW Turkic) and Eastern Armenian (Indo-European). Persian, Azerbaijanian, and Eastern Armenian are all spoken in Iranin Azerbaijan.

Most other languages spoken in that region are NW Iranian. These languages show evidence of Feature 1, except Kurmanji, Sorani, and Vafsi, which are located on the western and southern periphery of the region. Evidence regarding the other features is currently lacking.

Analysis of an Old Anatolian Turkish manuscript shows that the pattern was present in SW Turkic as early as the 14th century (predating written evidence of a distinct Azerbaijanian language). The fact that the Persian text from which this was translated lacks the pattern indicates that the pattern was not a translation effect.

Analysis of an Early New Persian text (composed in the 11th and copied in the 14th centuries) shows the pattern for human objects but not for inanimates. If these texts are representative, one possibility that suggests itself is that Turkic influence may have eliminated the animacy distinction in Persian. This influence may date from the Safavid state (founded ca. 150), the rulers of which were Persianized Turks who spoke a variety of Middle Azerbaijanian that might actually have been a mixed language incorporating Ottoman elements (Stein 2005: 228).

Much work remains to be done to establish the validity of the proposed isogloss, but the preliminary results are supportive of the hypothesis.

References


Differential Object Marking in Turkic and Persian as a Contact Phenomenon


Greg Key
UA School of Middle Eastern & North African Studies
PO BOX 2101888
Tucson, AZ 85721-0158

gkey@email.arizona.edu
Figure 1: Languages demonstrated to exhibit Features 1, 2, and 3. Azerbaijani (SW Turkic), Eastern Armenian (Indo-European), Persian (SW Iranian)

(Basic map taken from http://www.world-geographics.com/maps/middle-east/map-of-iran/)
Figure 2: NW Iranian languages. Underlined languages show evidence of Feature 1; languages not underlined clearly lack it.

(Basic map taken from http://www.world-geographics.com/maps/middle-east/map-of-iran/)