The Unstressed -i in Written Persian Discourse
Laura Crain

1. Background

In recent years many studies have been carried out on
the grammatical opposition of definiteness and
indefiniteness, focusing both on the theory of definiteness
(Givon 1977, Chafe 1976, Du Bois 1980) and on how
definiteness distinctions are realized in particular
languages (Comrie 1979, Schuh 1977, Du Bois 1980). Many of
these studies were conducted using a discourse approach,
examining the construction or opposition under consideration
in the context of entire texts rather than in isolated
sentences or groups of sentences. This approach has proven
to be invaluable in that definiteness marking is generally
dependent on identifiability, which in turn is dependent on
the entire discourse, all that has gone before a particular
mention of an entity and all that is to follow (Givon 1983,
Hopper and Thompson 1984).

The present study contributes to the understanding of
definiteness in discourse by examining a particular marker
in a particular language, the nominal marker -i in Persian.
-i has often been described as an indefinite marker, and
this paper will discover whether this is a sufficient and
complete description of its function. In so doing we will see how grammatical and discourse constraints combine to account for the distribution of -i. Furthermore, we will gain insight into how one function of a morpheme can be extended to encompass other functions which are related but not equivalent.

In this paper I show that the marker -i appears on three different classes of NP's which are grammatically and pragmatically diverse, and different discourse motivations account for the different classes. However, all of the environments where -i occurs can be understood in relation to its central function of marking non-identifiable new mentions.¹

The relationship between the different functions of -i will be considered further in Section 7. Sections 1 and 2 are devoted to a description of the definitions and methods used in this study. Sections 3 through 6 report the results of the investigation, and the analysis and conclusion are given in section 7.

1.1 Question

A noun phrase in Persian can be marked in three ways which relate to its grammatical role and/or its identifiability. The following are the possible markings on a noun phrase:
1) -ra (for "definite direct objects")
2) -i (for "indefinites")
3) (for all other NP's)

My goal in this paper is to examine the possible functions of -i by looking at it in discourse. As just mentioned, -i alternates with ra and , so clearly the results of this investigation will have implications for the description of the other types of nominal marking. Because of the large number of homophonous -i morphemes in Persian, the -i being investigated in this paper is distinguished by calling it the unstressed -i.²

2. DATA and METHODS

Four written texts with a total of 1,051 clauses were used in this study. Two of the texts are folk tales written in third person, and the other two are short stories, one in first person and the other in third person. In many of the tables below, the results are tabulated according to text, so the initials A, M, P, and Q represent the four texts used in this study. In addition, each of the examples in this
paper is taken from one of the four texts, unless otherwise indicated.

The method of this study is to pick out all of the NP's with -i in the corpus. For each NP with -i various grammatical and pragmatic information is recorded, such as its grammatical role, discourse identifiability, shape (noun or pronoun), composition (N + A, N + N, bare N), and its verb (negative, subjunctive, etc.). Examining this information, the next step is to make some generalizations about where -i occurs in the data. Any of the factors could be related to the appearance of -i. Does it appear only on subjects? Only on non-identifiable NP's? Only with subjunctive verbs? Only on an NP composed of N plus adjective? If any generalizations can be made, one then has to investigate the converse prediction. If we discover that -i is generally associated with a particular role, we then have to find out if that particular role is generally associated with -i. For example, to determine if -i is always and only a marker of subjects, we have to ascertain whether -i is required on all subjects and whether subjecthood is required of all -i's. This procedure is supplemented by a native speaker consultant, who is consulted about meaning differences between sentences and who provides further examples of the construction under discussion.

In undertaking such a study, certain terms need to be defined. These will be discussed below.

2.1 Grammatical Definitions

In this analysis, a clause is identified as a predication containing a single finite sentence-final verb. Modal verbs, which are identified by their subjunctive main verb complements, inflect for person and number but generally do not occur sentence-finally, so they are meant to be excluded from defining a clause.3

(1) na-tavānest-am nesbat be ān bitafāvatot (A79)
   NEG-can.PST-1SG regarding to that indifferent
   be-mēn-am
   SJ-remain-1SG

'I couldn't remain indifferent to that'

(2) mi-xāh-am ān-rē be-bus-am (P200)
   IMP-want.PRES-1SG that-RA SJ-kiss.PRES-1SG

'I want to kiss it'
The sentences in examples (1) and (2) consist of one clause each, although they each contain two inflected verbs.

A **subject** is defined as the NP with which the verb agrees in person and number. The subject may be either overt or non-overt, i.e. it may appear as a lexical NP or pronoun, or it may not appear at all ("zero subject"). In this study the distinction between the subject of an intransitive clause and the subject of a transitive clause is significant only in Section 7, where the grammatical roles of new mentions are discussed. Up to that point, **subject** will be used to refer to subjects of both transitive and intransitive clauses.

**Object** is identified as the NP that can take -rā. This was determined with the help of the consultant. Although most non-direct arguments in Persian are accompanied by a preposition, in certain contexts the preposition can be omitted. But these NP's, which can never take -rā, are still counted as obliques rather than objects. In the following examples we see an oblique without a preposition (3a), the equivalent sentence with a preposition (3b), and the ungrammatical sentence with -rā (3c).

(3a) vali mī-rav-ad mādrese
but IMP-go.PRES-3SG school
'M but she goes to school'

(3b) vali mī-rav-ad be mādrese/ vali be mādrese
but IMP-go.PRES-3SG to school miravad
'M she goes to school'

(3c) *vali miravad mādrese-rā / *vali mādrese-rā miravad

The other grammatical categories used in this study besides subject and object are oblique, predicate nominal, and incorporated noun. An oblique NP in this study is an NP which is the object of a preposition, a genitive NP, or a time or place adverbial (place adverbials are usually accompanied by a preposition, time adverbials usually are not). A **predicate nominal** is identified as a non-subject argument in a clause with a copula or copula-like verb. In the following examples, the underlined NP is counted as a predicate nominal.
Incorporated nouns are identified and discussed further in Section 1.7.

2.2 Discourse Definitions

Non-identifiable information is that for which the speaker/writer assumes the hearer/reader will not be able to pick out and establish reference, provided information already available within the universe of discourse (Chafe 1976, Payne 1984). The following examples, taken from the beginning of two stories, illustrate non-identifiable NP’s.

(7) pedar va mādar-i bud-and ke doxtar-i (P1)
father and mother-I be.PST-3PL that daughter-I
dāšt-and
have.PST-3PL

'There was a father and mother who had a daughter'

(8) dar madrese moallem-i dāšt-im
in school teacher-I have.PST-1PL

'In school we had a teacher...'

Identifiable information is that for which the speaker/writer assumes the hearer/reader can pick out and establish reference, provided information already available within the universe of discourse (Chafe 1976, Payne 1984). Identifiable NP’s in the next two examples are underlined.

(9) az in šahr be šahr-e digar-i be-rav-im (Q10)
from this city to city-LK other-I SJ-go-1PL

'Let’s go from this city to another city'
Referential NP's are those which are used to speak about an entity as existing and bounded within the universe of discourse, with continuous identity over time (Du Bois 1980, Payne 1984). In the next two examples, non-referential NP's are shown. In the first example, korsi is non-referential because it is used just as an adjective, and has no independent identity in the discourse. In the next example, zan-e gong is non-referential because it is in the scope of a negative, and does not exist in the universe of discourse.

Individuated NP's are those which are distinguished from their own referent class and from their own background (Hopper and Thompson 1980). Individuation is a relative concept, and a particular NP cannot be classified as absolutely individuated or non-individuated. In the following examples a few relatively non-individuated NP's are illustrated. Individuation will be discussed further in section 7.
The old woman picked up some pieces of clothing and also some knock-out medicine.

And her teacher had given her a lot of hope.

Differing from the traditional grammatical terms definite and indefinite, which usually denote the grammatical marking of an NP, identifiable and non-identifiable refer to the pragmatic status of an entity in the discourse, which is what I wish to refer to here.

2.3 Grammatical Constraints on -i

-i occurs in two positions in the NP, attached to the head noun or suffixed to the last word in the NP. These two positions are illustrated in the following examples.

'Suddenly he noticed that a beautiful girl came out of it.'
The former position, where -i is suffixed to the head noun, is acceptable for NP's composed of N + A, however it is generally excluded for genitive constructions, i.e. N + N NP's. The compound verb construction is another place where we see the interaction of -i with grammatical constraints. This will be discussed in the following section.

2.4 Compound Verbs

A construction in Persian which interacts with the occurrence of -i on an NP is the compound verb or noun incorporation construction. Compound verbs are made up of two constituents: the first can be an adjective, preposition, particle (having no independent syntactic status), noun, verbal noun, or prepositional phrase, and the second is a verb. In this discussion I focus on those compound verbs whose first element is a noun, since they are the ones that stand in opposition to an NP with -i.

Noun incorporation is a process whereby a noun stem is combined with a verb to form a new predicate denoting a unitary concept. The noun loses its semantic and syntactic salience, and serves to narrow the scope of the verb rather than refer to a specific entity.

The marker -i can appear on a noun in a compound verb construction, but when this happens the phrase is generally no longer considered a compound verb. A compound denotes a unitary concept; as Mithun (1984) discusses at length, specific, individuated entities are incompatible with noun incorporation. When an -i is added to an incorporated noun, the noun becomes individuated in a way that a bare noun is not. Adding an -i to a compound verb can change the meaning of the predicate, change the case relations within the sentence, or change the aspect of the verb.

Examples (17) and (18) illustrate a change in case relations. In (17), vasle 'patch' is an incorporated noun and has no grammatical case role, and lebēs is the direct object. In (18), in contrast, vasle-i is the direct object.
(17) mādar-e man lebās-e ma-rā vasle zad (elicited)
    mother-LK I dress-LK I-RA patch hit.PST.3SG
    'My mother patched my dress'

(18) bā forūš-e ānhā vasle-i be šekam-e xānevāde-aš
    with sale-LK they patch-I to stomach-LK family-his
    mi-zad
    IMP-hit.PST.3SG
    'With their sale he would put a patch on his family’s stomach'

In the next examples there is a change in both meaning
and case relations.

(19) sib zamini xeyli dust dār-am (elicited)
    apple earth very friend have.PRES-1SG
    'I really like potatoes'

(20) dar ālmān dust-i dāšt-am,
    in Germany friend-I have.PST-1SG
    ammā u digar dar ānjā nist
    but s/he any more in there NEG.be.PRES.3SG
    'I had a friend in Germany, but she isn’t there any more'

Finally, examples (21) and (22) illustrate a change in
the aspect of the verb. (21), without -i, is an ongoing
activity. When the -i is added, the noun becomes "singular"
and indicates that the action of the verb occurs only once,
i.e. the predicate takes on a punctual interpretation.

(21) barādar-aš atse zad (elicited)
    brother-her sneeze hit.PST.3SG
    'Her brother sneezed (=sneeze-hit)'

(22) barādar-aš atse-i zad (Q40)
    brother-her sneeze-I hit.PST.3SG
    'Her brother gave a sneeze'

In this study I have counted unmarked NP’s appearing
immediately adjacent to the verb as incorporated nouns. A
nominal constituent of a compound verb marked by -i,
however, I have counted not as an incorporated noun, but
according to its independent grammatical role, subject, object, or oblique.

3. DISCUSSION

3.1 Results

The number of full NP's (i.e. overt lexical nouns, excluding pronouns) was counted for each text, and the number of those NP's with -i. The totals are shown in Table 1.

<table>
<thead>
<tr>
<th>Clauses</th>
<th>Full NP's</th>
<th>NP's w/ -i</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>132</td>
<td>171</td>
</tr>
<tr>
<td>M.</td>
<td>427</td>
<td>730</td>
</tr>
<tr>
<td>P.</td>
<td>208</td>
<td>283</td>
</tr>
<tr>
<td>Q.</td>
<td>284</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>1,051</td>
<td>1,486</td>
</tr>
</tbody>
</table>

Table 1. Count of Clauses and Full NP's in 4 texts (percentages computed by row)

Averaging over all the texts, 10% of the full NP's have -i. The contexts in which these NP's with -i occur will be discussed in the following sections.

3.2 Distribution of NP's with -i

After examining the 143 occurrences of -i in the text, it became clear that there are three different environments, defined according to both grammatical and pragmatic factors, where -i consistently appears. These environments are (i) a non-identifiable first mention; (ii) an NP in the scope of a negative quantifier or verb; and (iii) a predicate nominal. Table 2 shows the number of -i's in each of these environments.
Table 2. Count of functions of NP's with -i (percentages computed by row)

Two additional categories, other than those just mentioned, were used in the count. "Other non-referentials" are those which appear in metaphors and similes or as a fixed expression in a compound verb (example 24). A few examples are given below.

(23) u ne-mi-dānest ke havādes hamcon (M77)  
    she NEG-IMP-know.PST.3SG that events like

zanjir-i u-rā az kudaki tā piri  
    chains-I she-RA from childhood to old.age

mi-kešān-ad  
    IMP-draw.PRES-3S

'She didn’t know that events like chains would draw her from childhood to old age'
The last category in Table 2, "other", includes all the occurrences of -i not covered by another category.

Looking at the bottom row in Table 2, where the totals are tabulated, we see that 70%, or over two-thirds of all NP's with -i, have the status of non-identifiable first mention. This might suggest that -i has something to do with information flow, roughly the way information is packaged in a discourse in order to convey its status to the reader/hearer. This will be discussed further in Section 7. At this point I would like to present a more detailed discussion of -i in each of the environments listed above.

4. Non-identifiable first mentions

This category covers all instances of -i on an NP that is introducing a referent into the discourse for the first time. In this use, -i appears only on the very first mention of a referent, and is not used to reintroduce the same referent later in the same text. At the same time, -i does not serve to introduce a whole referent class, but only an individual. That is, if a well is introduced at some point in the discourse, and at some later point a different well is mentioned, both mentions could be marked with -i. In (25) the two mentions of different referents in the same referent class appear in the same sentence, and both are marked by -i.

(25) va otobus dobâre piš mi-raft az (M35) and bus again forward IMP-go.PST.3SG from rustâ-ı; be rustâ-ı; village-I to village-I

'And the bus proceeded from (one) village to (another) village'

Although rustâ, 'village' has already been introduced in the clause, the immediately following instance of rustâ, is
marked with -i because it is not the same village but another village which is being mentioned -- the referent is different.

Another point that deserves comment is that some new mentions are not marked by -i: identifiable NP's and incorporated nouns. Identifiable first mentions, such as proper names, universally known entities (e.g. 'sun'), and culturally assumed entities (e.g. 'prince'), cannot take -i. Also incorporated nouns do not take -i because -i gives the noun an independent status apart from the verb. Even when an incorporated noun is mentioned for the first time, it does not need to be introduced as a new participant, because it serves merely to narrow the scope of the verb.

5. Negative Sentences

Again referring back to Table 2, we see that 18% of the NP's with -i in this corpus appear within the scope of a negative. It is by no means necessary that the NP's in a negative sentence appear with -i. Table 3 below shows the total number of negative sentences in the data and the percentage that contain at least one NP with -i.3

<table>
<thead>
<tr>
<th># of neg S's</th>
<th># of neg S's containing -i</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21</td>
</tr>
<tr>
<td>M</td>
<td>45</td>
</tr>
<tr>
<td>P</td>
<td>14</td>
</tr>
<tr>
<td>Q</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 3. Negative sentences

Givon (1979) discusses referentiality in negation at length. He points out that, although languages generally allow referential-indefinite objects in affirmative sentences, in negative sentences most languages either allow them only infrequently or exclude them altogether. The same is true of subjects: referential-indefinite subjects in negative clauses are disfavored cross-linguistically.
The same restrictions apply in Persian: the subjects and objects of negative clauses are either referential-definite or non-referential. In the data for this study, no cases of referential-indefinites in negative clauses were found.

In the following examples I show negative sentences with identifiable NP's (26), negative sentences with non-referential NP's without -i (27), and some with non-referential NP's with -i (28).

(26) a. ammā be vāqe' fārsi ne-mi-dān-im (A9)
   but to reality Persian NEG-IMP-know.PRES-1PL
   'But we really don't know Persian'

   b. va nowzūdī-rā na-did
   and newborns-RA NEG-see.PST.3SG
   'And he didn't see the newborns'

(27) a. man zan-e gong ne-mi-xāh-am (P206)
   I wife-LK mute NEG-IMP-want.PRES-1SG
   'I don't want a mute wife'

   b. xāhar o barādar-i bud-and ke pedar o
   sister and brother-I be.PST-3PL that father and
   mādar na-dāšt-and
   mother NEG-have.PST-3PL
   'There was a sister and brother who didn't have
   father and mother'

(28) a. vali az doxtaru asar-i na-did (P139)
   but from girl trace-I NEG-see.PST.3SG
   'But he didn't see a trace of the girl'

   b. digar ciz-i dar xāne na-bud (P65)
   any.more thing-I in house NEG-be.PST.3SG
   'There wasn't a single thing left in the house'

These examples illustrate the fact that -i cannot appear on an identifiable NP in the scope of a negative, and non-referential NP's in a negative sentence sometimes take -i and sometimes do not. The first finding is expected: -i seems not to be associated with identifiable NP's in any context. But how can we explain the difference between the
contexts where \(-i\) does and does not occur in the scope of a
negative?

There are two complementary explanations for this use
of \(-i\). One is that \(-i\) carries the sense of 'one' here. The
combination of 'not' + 'one' is widely attested in different
languages as an independent negative marker. Examples from
two Indo-European languages are French aucun(e) '(not) any,
(not) one' from un(e) 'one' and Icelandic einginn 'none'
from ein 'one' and gi 'not'. Although the construction in
Persian has not coalesced to form a single word or morpheme,
it's force is the same as in French or Icelandic: it shifts
the focus of negation from the verb to the noun. The
difference between negating the verb and negating the noun
is shown in the following English examples.

\[(29) I \text{ didn't see a book/any books.}\]
\[(30) I \text{ saw no books.}\]

I am not claiming that the Persian cases with \(-i\) are the
exact counterparts of the English examples, but only that in
negative sentences with \(-i\) the negation is centered on the
noun, whereas those without \(-i\) have verb-centered negation.

One further constraint governs the use of \(-i\) with
negatives, and that was discussed earlier, in the section on
\(-i\) and yek (section 1.4). Bare nouns in object position
tend not to occur in negative sentences for the same reason
they do not occur in affirmative sentences: a bare N + verb
is construed as a compound verb. If the object noun in a
negative sentence does not form a unitary concept with
the verb (cf. Mithun 1984), lack of marking is dispreferred,
and the noun will generally take \(-i\). The following table
shows the number of non-identifiable NP's in negative
sentences appearing in various grammatical roles, and
whether they have \(-i\) or not.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>S</th>
<th>O</th>
<th>IN</th>
<th>Obl</th>
<th>PN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP's with (-i)</td>
<td>1</td>
<td>3</td>
<td>14</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>NP's without (-i)</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Non-identifiable NP's in negative sentences
The table shows that most non-identifiable NP's in object position appear with -i. And in fact those which occur without -i are understood as incorporated predicates, even though they are not lexicalized compound verbs.

The second explanation for the appearance of -i in negative sentences is applicable especially to mass nouns. Horn (1978) discusses "negative strengtheners": morphemes or particles which are added to existing negative adverbs for a strengthening effect (148). After discussing examples from French, Germanic languages, and Spanish, he goes on: "But the most universally consistent class of such strengtheners is a set of words referring to small portions, the perceptually or psychologically indivisible atoms of any category. This set includes such items as ENGLISH (not a bit/scrap/jot/iota/shred/soul)....These classes of expressions are productive in ENGLISH and other languages" (1978:149). Although we are not talking about a class of expressions in Persian but a single expression, the -i has the same effect of referring to small portions. This is clear with mass nouns which, with -i attached, are understood as 'not a bit/not a drop', etc. The following examples show the contrast between mass nouns with and without -i.

(31) mā ham āb na-dār-im (Q24)
we also water NEG-have.PRES-1PL
'We don't have water either'

(32) mā ham āb-i na-dār-im (elicited)
we also water-I NEG-have.PRES-1PL
'We don't have a drop/a bit of water either'

(33) va pul-i ham barāye xaridan na-dār-im (M354)
and money-I also for buying NEG-have.PRES-1PL
'And we also don't have a cent to buy anything'

For mass nouns -i picks out the "psychologically indivisible atom" of the category, and likewise for count nouns: one individual is the smallest unit of the category. Hincha makes this argument in his 1961 paper: "With stems that designate masses, the stem form with /-i/ designates a part of the continuum.... If the designated, on the other hand, is composed of individual objects, then the form with /-i/ refers to the smallest part that the type can
represent, down to an individual object" (168). So the use of -i in negative sentences can be seen as an instance of 'one' + NEGATIVE, or as an example of what Horn calls negative strengtheners.

6. Predicate Nominals

In Table 2 we saw that only 5% of the NP's with -i are predicate nominals. However, 19% (7/37) of all predicate nominals are marked with -i, so it seems that -i marking may have some significance for the class of predicate nominals.

Table 5 shows all the predicate nominals which occurred in the data tabulated according to whether they are a bare N (with no modifiers), an N with nominal modifiers (N + N), or a noun modified by an adjective (N + A). Each category is subdivided into those that appear with -i and those that do not.

<table>
<thead>
<tr>
<th></th>
<th>bare N</th>
<th>N + N</th>
<th>N + A</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN's w/-i</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>7 (19%)</td>
</tr>
<tr>
<td>PN's w/o -i</td>
<td>12</td>
<td>16</td>
<td>2</td>
<td>30 (81%)</td>
</tr>
<tr>
<td>Total PN's</td>
<td>14</td>
<td>16</td>
<td>7</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 5. Composition of Predicate Nominals in corpus

There seems to be a split among predicate nominals as to where -i is likely to occur. -i is more likely to occur on an N + A predicate nominal, less likely to occur on a bare N predicate nominal, and unlikely to occur on an N + N predicate nominal.

The lack of -i marking on N + N predicate nominals can be easily explained. -i cannot break up a genitive, and generally -i cannot appear on the second N of an N + N construction because it would have a semantic effect on the wrong N. For example, the phrase moallem-e qor'an-i (teacher-LK Qor'an-I) is unacceptable because it conveys the unfelicitous meaning 'teacher of a Qor'an' rather than 'a Qor'an teacher'. Similarly the phrase mobser-e kelîs-i (monitor-LK class-I) sounds wrong to a native speaker.
because it would mean 'monitor of a class' rather than 'a class monitor'. It is not impossible, however, to find an -i with an N + N construction. In the corpus seven instances of -i in an N + N construction appear, some cases with -i occurring on the first noun and some with -i on the second noun. The cases where -i appears on the first noun are adjectival and appositional constructions, and those where -i appears on the second noun are partitive or -- in one example -- the -i is intended to qualify the second noun rather than the first. That example is shown below.

(34) va moallemaš be u besiār omid dād-e (M46) and teacher.her to she much hope give-PTCP

bud omid-e ḳyande-i roḵan
be.PST.3SG hope-LK future-I bright

'And her teacher had given her a lot of hope, hope of a bright future'

Another factor which makes it unlikely for -i to occur on N + N constructions is that many of them are identifiable. Given the other uses of -i, it is not surprising that it also does not appear on identifiable predicate nominals. Example (35) shows an identifiable N + N predicate nominal.

(35) in dars xānān-e ma'sume ham šod-e (M275) this studying-LK ma'sume also become-PTCP

balā ye jān-e mā
pain-LK life-LK we

'This studying of Masume's has become our life's pain.'

Turning to the other two categories, bare N and N + A predicate nominals, a different sort of explanation is in order. I would like to suggest that bare N predicate nominals generally do not get marked by -i because of their propensity for becoming incorporated nouns. As discussed in Section 1.7, noun incorporation is a productive process in Persian. Many N + V combinations, although they are not lexical entries, may be considered compound verbs when they
are used to denote a unitary activity or state, with the noun out of focus as an individual entity. In the case of predicate nominals, a bare N + budan 'to be' is readily seen as a compound verb because the verb adds so little semantic content. Thus, while a concatenation like zan sostan (woman to.wash) 'to woman-wash' might not be easily accepted as an institutionalized activity or generalized predicate, zan budan 'to woman-be' can be seen as a unitary concept: it does nothing more to the concept woman than predicate it of the subject. So almost any bare N in the company of budan can be considered an incorporated noun.

Mithun (1984:863) makes an observation related to this point: "Certain types of V's are more likely to incorporate than others. Several significant factors enter into this likelihood. [One] significant factor is the scope of the V. General V's, which take much of their meaning from their arguments, are more likely to incorporate than those with narrow scope; thus 'to be good' or 'to have' incorporate especially often." This holds true for Persian. The most common verb appearing in compounds is kardan 'to do', which is wide in scope. I am making a similar point for budan: it has almost no meaning apart from its argument, and thus is a good candidate for compounding.

N + A combinations, on the other hand, cannot become incorporated nouns. The -i on N + A predicate nominals signals their independent status: they are entities which are not joined with the verb. They still have an individual semantic role. An N + A predicate nominal is more specified than a bare N predicate nominal, and the -i serves to keep it distinct from the verb.

This distinction between N + A and bare N predicate nominals has become grammaticized to a certain extent. It is not necessarily true that N + A combinations always have more semantic content and are less able cross-linguistically to become incorporated nouns than bare N's. A predicate like 'to be a good person', as Mithun points out (1984:863), is a common type of compound cross-linguistically because it is highly generic. But in Persian such a phrase would be rendered with -i, i.e. xad-e xub-i budan (person-LK good-I to.be). N + A combinations cannot become incorporated nouns in Persian, so they almost always get marked with -i when they appear in a clause with budan. And since an unmodified N + verb combination is the typical composition of a compound verb, when the verb is budan, which is very general in meaning, the N usually appears as an incorporated noun.
7. Analysis

Table 6 below shows all the full NP's in the data tabulated according to grammatical role, and Table 7 shows the corresponding counts for NP's with -i. In this section I use A for the subject of a transitive clause, S for the subject of an intransitive clause, and O for the object of a transitive clause.

**Table 6. All Full NP's (percentages computed by row)**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>S</th>
<th>O</th>
<th>Incorp</th>
<th>Oblig</th>
<th>PN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Noun</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4</td>
<td>22</td>
<td>33</td>
<td>16</td>
<td>81</td>
<td>15</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>(2%)</td>
<td>(13%)</td>
<td>(19%)</td>
<td>(9%)</td>
<td>(47%)</td>
<td>(9%)</td>
<td>100%</td>
</tr>
<tr>
<td>M</td>
<td>37</td>
<td>130</td>
<td>142</td>
<td>54</td>
<td>352</td>
<td>15</td>
<td>730</td>
</tr>
<tr>
<td></td>
<td>(5%)</td>
<td>(18%)</td>
<td>(19%)</td>
<td>(7%)</td>
<td>(48%)</td>
<td>(2%)</td>
<td>100%</td>
</tr>
<tr>
<td>P</td>
<td>15</td>
<td>56</td>
<td>37</td>
<td>36</td>
<td>134</td>
<td>5</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>(5%)</td>
<td>(20%)</td>
<td>(13%)</td>
<td>(13%)</td>
<td>(47%)</td>
<td>(2%)</td>
<td>100%</td>
</tr>
<tr>
<td>Q</td>
<td>21</td>
<td>81</td>
<td>51</td>
<td>32</td>
<td>115</td>
<td>2</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>(7%)</td>
<td>(27%)</td>
<td>(17%)</td>
<td>(11%)</td>
<td>(38%)</td>
<td>(1%)</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>77</td>
<td>289</td>
<td>263</td>
<td>138</td>
<td>682</td>
<td>37</td>
<td>1486</td>
</tr>
<tr>
<td></td>
<td>(5.2%)</td>
<td>(19.4%)</td>
<td>(17.7%)</td>
<td>(9.3%)</td>
<td>(45.9%)</td>
<td>(2.5%)</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 7. Full NP's with -i
Looking at Table 7, we see that an NP with -i is most likely to show up in the O or Oblique role, next most likely to show up as an S, and not very likely to show up in the roles of A, Incorporated Noun, or Predicate Nominal. These results agree in a general way with those obtained by Du Bois (1984) for new mentions in Sacapultec. In this paper he supported the claim that new mentions show a preference for the S and O roles and various Oblique roles. Du Bois’ Figure 2 is reproduced here along with the totals of Table 7 shown in the form of a bar graph in Figure 1.

I am not suggesting that -i can be simply equated with 'new mention', since it has other functions, but it is interesting to notice the similarity in the two sets of results. The most noticeable difference between Du Bois’ results for Sacapultec and those reported here for Persian is in the difference between the percentage of new mentions (or NP's with -i) in the S role and in the O role. This is due in part to the fact that I have used percentage of NP's with -i rather than percentage of new mentions. And in fact when the negatives are factored out, the percentages are much closer: S -- 22%, O -- 36%, Obl -- 32%. But there is
still a 14% gap between the S and O roles, which can be explained by a closer examination of the types of verbs that introduce new entities in the S position.

Among the 25 tokens of S's with -i recorded in Table 7, fifteen different verbs are represented. It turns out that these fifteen verbs are all either stative or motion verbs. Examples are budan 'to be', peyda šoden 'to be found', gozaštan 'to pass (of time)', dar Šamadan 'to come out', and raftan 'to go'. The intransitive verbs in the data which serve to introduce their subject are limited to a few semantic classes, although the overall semantic range of intransitive verbs is much wider. Thus this contributes to limiting the number of new mentions in the S position: they can only appear with a small subset of intransitive verbs.

We have found that -i fulfills different functions according to the type of sentence in which it occurs: it can serve to mark a new mention, it can act as a negative strengthener, or it can give an NP individual status to show it is not an incorporated noun. This is not necessarily a bad result: although these functions are different, they are not at odds with each other. As Du Bois (1980:207-8) explains for the English article system, a contrast made in one part of the grammar may be transferred to another part of the grammar to mark a different contrast which is, nonetheless, not incompatible with the first. Persian is using -i as a marker of non-identifiable new mentions, and when it wishes to strengthen a negative by designating the smallest unit of the referent, -i is also employed. In both cases -i serves to call attention to an entity in the discourse -- either to its entrance into the discourse or to its utter non-existence in the discourse (cf. Ex. 28b, 'there wasn't a single thing left'). The use of -i in negative sentences is very like a first mention: it can be thought of as introducing a new referent by asserting its non-existence.

Lastly, let us see if there is a connection between the central function of -i and its role in marking predicate nominals. Assuming that the notion of individuation has something to do with -i's use on predicate nominals, is there a connection between non-identifiable first mentions and individuation?

In their 1980 paper on transitivity, Hopper and Thompson give six properties which correlate with individuation (253). These are proper vs. common; human and/or animate vs. inanimate; concrete vs. abstract; singular vs. plural; count vs. mass; and referential and definite vs. non-referential. In each pair the first property corresponds to high individuation, the second to
low individuation. In order to find out if there is a connection between NP's with -i and high individuation, I calculated an individuation rating for every non-identifiable entity introduced into the discourse: these included NP's with -i and incorporated nouns. A subset of the properties listed above was used in rating these entities. Proper vs. common was not used, as non-identifiables cannot be proper nouns, and definite was ignored because I limited my sample to indefinite nouns. For the remaining five properties, one point was given to each entity for every property of high individuation that it possessed.

The results of the individuation count are given in Table 8.

<table>
<thead>
<tr>
<th></th>
<th>NP's with -i</th>
<th>NP's w/o -i (IN's)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.3</td>
<td>2.6</td>
</tr>
<tr>
<td>M</td>
<td>3.4</td>
<td>2.7</td>
</tr>
<tr>
<td>P</td>
<td>3.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Q</td>
<td>3.9</td>
<td>3.0</td>
</tr>
<tr>
<td>Overall</td>
<td>3.5</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 8. Ave. Individuation Ratings for non-identifiable NP's

The table shows that, although the difference is not great, the pattern is consistent: for each data source as well as overall, NP's introduced into the discourse with -i are more individuated than those that enter without -i.

This finding supports the claim I am making about predicate nominals. When a noun in the position of a predicate nominal is accompanied by an adjective, or when it cannot form an incorporated predicate with the verb for other reasons, it is marked by -i to signal its individuated status vis-à-vis noun incorporation: a noun phrase with -i affixed cannot be considered part of a compound verb. This use of -i relates to its use on first mentions. In both cases those NP's accompanied by -i are more individuated than those which are not.
7.1 Special Cases

Before I summarize the findings in this study I would like to say a few words about some of the special cases. There are several examples of NP’s with -i that are not typical cases of non-identifiable first mentions, but they have been counted as such nevertheless. I will now explain each case in turn. Three instances of -i in this category are not strictly speaking first mentions but can still be seen as non-identifiable from the point of view of a hearer/reader in the story. The examples are shown below.

(36) yek daf’e motavaje šod ke (P86)
one time noticing become.PST.3SG that

doxtar-i zibā az ān birun āmad
girl-I beautiful from that out come.PST.3SG

'Suddenly he noticed (=noticing-become) that a beautiful girl came out of it’

(37) Šāhzāde negāh-e bālā kard. (Q61)
prince look-LK up do.PST.3SG

cāšmaš be doxtar-i oftād
eye.his to girl-I fall.PST.3SG

'The prince looked up. His eyes fell on a girl’

(38) barādar-i dār-am (Q140)
brother-I have.PRES-1SG
‘I have a brother’

In each of these cases the NP with -i which seems to be introduced has in fact been mentioned in the text several times previously. So strictly speaking these -i’s are not coding first mentions, nor are the NP’s non-identifiable. However in each case the perspective has shifted so that the -i actually indicates non-identifiable information from the perspective of one of the characters in the story. In (36) the prince (a zero subject in this sentence) is watching and sees a girl he has never seen before. Although she has been a principal character in the story up to this point and is identifiable to the reader, she is coded by an NP + -i in accordance with the perspective of the prince. (37) is
exactly parallel to (36): the prince (in another story) sees a girl who is identifiable to the reader. In (38) the girl is talking to the prince whom she has just married, and who doesn't know she has a brother. Since it is conversation, it is natural for her, the speaker, to take the hearer's point of view.

It is clear why an -i is used in these contexts: the writer wishes to show the perspective of one of the characters, and so uses a marker of non-identifiable first mention.

Up to this point, we have seen -i as a mark of non-identifiability from the reader/hearer's point of view. This is the first time we find -i marking non-identifiability from someone else's perspective. But the use of -i in these special cases fits very well with our original hypothesis: it is still, in spite of the shift in perspective, a marker of non-identifiability.

7.2 CONCLUSION

To summarize the findings of this paper, we have seen that the unstressed -i in its central function relates to information flow in discourse. Specifically, it marks non-identifiable NP's. The other functions of -i, while not collapsible into one over-arching 'meaning', are all related to the central function of -i, like satellites of a central core. In this way negative strengthening, information status, and individuation can all be seen as different aspects of -i's function in marking non-identifiable NP's in discourse.
Identifiable (as opposed to non-identifiable) new mentions are proper names and culturally identifiable entities such as 'the president' in the U.S. or 'the prince' in fairy tale Iran. Other nouns, such as 'house' and 'school', are also introduced into a discourse as identifiables, because every member of a language community recognizes that a unique 'house' and a unique 'school' can be associated with every person. So for example, in the sentence She went in the house, house can be introduced as identifiable because the hearer interprets it as 'the house where she lives'. The cultural assumptions underlying the encoding of these entities as identifiable new mentions are that everyone lives in a house and everyone attends or has attended a school.

Most of the other -i morphemes retain final stress, which is usual for nouns and adjectives in Persian, while the -i here is never stressed. The only other -i without stress is sometimes claimed to be the same morpheme as the "indefinite" -i here; it is required on the head noun of a restrictive relative clause. Based on historical evidence and Windfuhr (1979), I maintain that these two -i's are separate morphemes, and will not consider restrictive relative clauses in this study.

Modal verbs can occur sentence-finally in elliptical constructions, as in the following example:

har kār-i kard ke doxtaru-rā biyār-e
every work-I do.PST.3SG that girl-RA SJ.bring-3SG

pā'īn na-tavēnest (Q63)
down NEG-can.PST.3SG

'No matter what he did to bring the girl down, he couldn't'

This type of example presents no problems to our clause-counting task: when a modal verb in a sentence appears without a main verb complement, the sentence is nonetheless counted as a clause.
Abbreviations: I = Persian unstressed -i; IMP = Imperfective; LK = linker (Persian grammatical formative); NEG = negative; PL = plural; PRES = present; PST = past; PTCP = participle; RA = Persian -rā (case marker); SG = singular; SJ = subjunctive; 1 = 1st person; 2 = 2nd person; 3 = 3rd person.

In fact none of the negative sentences in these texts contain more than one occurrence of -i.
REFERENCES


