Identifiability and verbal cross-referencing markers in Hungarian

Casper de Groot
ACLC – Universiteit van Amsterdam

Summary
Many languages employ articles as an expression device to mark identifiability. There are, however, other means to mark identifiability in languages. In this paper I show that one of the two sets of cross-referencing markers on the transitive verb in Hungarian serves this purpose, i.e. it represents the instruction from the speaker that he considers the referent of the object argument to be identifiable to the hearer. Next, attention is paid to four sets of Hungarian data which are seemingly at odds with this claim. These sets relate to the use of particular person distinctions, indefinite noun phrases which contain either a possessive restrictor or the interrogative pronoun melyik ‘which’, and to relative clauses. Apart from a descriptive account of the relevant data, the paper also offers a theoretical account within the framework of Functional Discourse Grammar, where the speaker assessment of identifiability of referents is formally represented by operators which apply to the Referential Act, represented by the variable R at the Interpersonal Level. Distinctions made at the Interpersonal Level thus affect the choice of verbal forms in Hungarian.

1. Preliminary remarks
This paper offers a detailed descriptive account of the use of two different sets of verbal cross-referencing markers in Hungarian. The data are of interest not only to descriptive linguists, but also to linguistic typologists and theoretical linguists. For this reason I will discuss the relevant distinctions against the background of a formal theory. Such a formal theory, I claim, must be able to accommodate components for interpersonal, semantic and morphosyntactic analyses and the interaction between them. Functional Discourse Grammar as developed by Hengeveld & Mackenzie (2006, 2008) is such a theory. The following notions from this theory are relevant to this paper:

(i) The Interpersonal Level deals with all the formal aspects of a linguistic unit that reflect its role in the interaction between Speaker and Addressee. At the Interpersonal Level we find Discourse Acts, which consist of an Illocution, the Speech Participants and a Communicated Content. Within the Communicated Content, the Speaker evokes one or more Subacts of Reference and Ascription.

(ii) The Representational Level deals with the semantic aspects of a linguistic unit.

(iii) The Morphosyntactic Level accounts for all the linear properties of a linguistic unit, both with respect to the structure of sentences, clauses and phrases and with respect to the internal structure of complex words.

I assume that this information is sufficient for the reader to understand the formalizations used. For further details the reader is referred to the introduction to this volume.

2. Identifiability
When uttering a message a speaker may refer to entities that he considers identifiable to the addressee (Hengeveld & Mackenzie 2008: 113-114). Such identifiability of referents may arise on the basis of different kinds of information: (i) the addressee may know the person or
thing in real life, e.g. Peter in I met Peter yesterday; (ii) the entity referred to may be visible to the addressee, e.g. that tree in I love that tree over there; (iii) the speaker may introduce an entity in discourse, after which the entity will be identifiable to the addressee, e.g. the man in the sequence I met a sailor last night. The man told me a nice story about a passage to India; or (iv) the addressee may identify entities through inference, e.g. the kitchen in It is a nice house, but the kitchen is too small (the presupposition being that every house has a kitchen).

Languages often have grammatical means to signal whether the addressee is to construe a referent, i.e. when the speaker considers the referent to be non-identifiable, or to retrieve a referent, when the speaker considers the referent to be identifiable. A marker typically used for this purpose is the definite article. There are, however, other markers too, as will be shown in this paper. By using such markers the speaker offers the addressee a clue, namely that the speaker assumes that the addressee will be able to identify the referent in one way or another. The addressee will act on this by trying to retrieve the appropriate referent.

3. Two sets of verbal cross-referencing markers
Hungarian has two sets of verbal cross-referencing markers. The first one is used as a default, whereas the second one is used on transitive verbs where the object is assumed to be identifiable (+id). The examples in (1) illustrate the different occurrences of the two sets. I refer to the sets as first form (1f) and second form (2f), respectively.

(1) a. A nők sétál-nak. [Intransitive]
   the women walk-3PL.1f
   ‘The women walk/are walking.’

b. Könyv-et olvas-ok. [Transitive, -id object]
   book-ACC read-1SG.1f
   ‘I read/am reading a book.’

c. Egy könyv-et olvas-ok. [Transitive, -id object]
   a book-ACC read-1SG.1f
   ‘I read/am reading a book.’

d. A könyv-et olvas-om. [Transitive, +id object]
   the book-ACC read-1SG.2f
   ‘I read/am reading the book.’

The two sets consistently apply to all person distinctions for all tense and mood distinctions, i.e. indicative, conditional, and imperative/adhortative mood (see appendix 1 for an example of all forms in a verbal paradigm). The distinction is also expressed on auxiliary verbs, as for instance in:

(2) a. Lát-ni fog-ok valami-t. See-INF FUT-1SG.1f something-ACC
   ‘I will see something.’

b. Lát-ni fog-om az-t. see-INF FUT-1SG.2f that-ACC
   ‘I will see that.’

The two sets are usually referred to as the ‘indefinite conjugation’ and the ‘definite conjugation’. The terminology of ‘(in)definite conjugation’ is, strictly speaking, not appropriate, since intransitive verbs, which do not have an object, also qualify for the indefinite conjugation as if there is an indefinite object. Hungarian grammars generally use
alanyi igeragozás ‘subjective conjugation’ and tárnyas igeragozás ‘objective conjugation’.

This characterization, however, is also inappropriate, since transitive verbs with an identifiable object agree not with just the object (+id), but also with the subject (person and number). Another opposition found is that of általános ‘general’ versus határozott ‘definite’. I opt for the use of a neutral terminology, i.e. ‘first form’ and ‘second form’ (De Groot 1989).

The second form applies when the referent specified by the object is identifiable. The referent may be a first order entity, i.e. an object that may be located in space, but also a higher order entity, such as an event, as illustrated in (3e).

(3) a. Lát-juk az-t.
    ‘We see that.’

b. (A) Mari-t lát-juk.
   ‘We see Mary.’

c. Lát-juk a magas fá-t.
   ‘We see the tall tree.’

d. Lát-juk az-t a magas fá-t.
   ‘We see that tall tree.’

e. Lát-juk, hogy esik.
   ‘We see that it is raining.’

In other cases, where the referent specified by the object is presented as not identifiable, the first form is used:

(4) a. (Egy) fá-t lát-unk.
    ‘We see a tree / trees.’

b. Lát-unk valami-t.
   ‘We see something.’

c. Lát-unk hat magas fá-t.
   ‘We see six tall trees.’

Objects of questions in which the Speaker inquires after a person, a thing, a quality or a number are also not identifiable. The Speaker does not know the referents, qualities or numbers, while assuming that the Addressee does.

(5) Ki-t / mi-t / milyen/hány fá-t lát?
   ‘Who / what / what kind of/how many tree(s) does (s)he see?’

The distinction between the two sets goes back to the very early stages of Hungarian. It is very likely that the second form in proto-Hungarian originates from a construction in which the object is both expressed by a pronoun and by a noun phrase in apposition to the
pronoun. The pronoun became a clitic in later times and then turned into an object agreement marker. This process can be represented schematically, in pseudo-English, as follows:4

(6)     Stage 1: Object noun phrase  Árpád kill the wild-hog
Stage 2: Object pronoun + apposition  Árpád kill it, the wild-hog
Stage 3: Object clitic + apposition  Árpád kill-it the wild-hog
Stage 4: Object agreement + noun phrase  Árpád kill-t the wild-hog

One example of genuine object marking has remained in present-day Hungarian, namely in those cases where the subject is specified for first person singular and the object for second person (singular or plural). The subject and object pronouns are optional, the marking on the verb obligatory:

(7)     (Én) látt-ak   (téged / titeket).
        I see-2Obj-1SG.Subj you.SG.ACC / you.PL.ACC
‘I see you.’

In Section 5 and 6 below I will discuss a number of seemingly incongruous uses of the 1f and 2f cross-reference markers. The question is whether these are genuine counterexamples to the analysis presented so far, or whether they may be explained through interaction with some other part of grammar. But before doing so, it is important to discuss several other domains in the grammar of Hungarian where identifiability plays a role.

4. The noun, possessives and the use of the article in Hungarian

The choice in cross-reference markers between 1f and 2f is not the only way to assess identifiability in Hungarian. For a good understanding of the apparent counterexamples presented in Section 5 and 6, let me first elaborate on two specific domains: set nouns and their occurrence in possessive constructions, and article use.

4.1. Hungarian set nouns

According to Rijkhoff (2002), Hungarian has nouns that denote sets of individuals, so-called set nouns. Rijkhoff shows that on the basis of morphosyntactic and semantic properties the ‘noun’ in the languages of the world can be divided into six sub-types or nominal subcategories, which he refers to as Seinsarten (‘modes of being’). Hungarian, a set noun type language, differs from e.g. English, which is a singular object noun type language. Differences arise for instance in the use of the plural and the use of the article. The use of bare nouns is found in set noun type languages more often than in singular object noun type languages, as can be seen in the following examples from Hungarian and English:

(8)     Hungarian                          English
a.  Van alma?   a.  Are there apples?
    COP.3SG apple
    ‘Are there apples?’
b.  Mari könyvet olvas.   b.  Mary is reading *(a) book.
    Mary book-ACC read.3SG.1f
    ‘Mary is reading a book.’
A set can have any cardinality: it may contain just one individual (singleton set (9a)) or it may consist of more individuals (multiple set (9b)). The marker -k in Hungarian (9c) is considered a collective marker rather than a plural marker (cf. Rijkhoff 2002: 51f).

(9) Hungarian \hspace{1cm} English
a. egy /‘eggy/ könyv \hspace{1cm} a. one book
one book
‘one book’
b. két könyv \hspace{1cm} b. two book*(s)
two book
‘two books’
c. könyv-ek \hspace{1cm} c. book-s
book-COLL
‘books’

In addition to Rijkhoff’s observations and analysis, I would like to claim that the Seinsart of a set may also apply to units larger than the noun. The nouns in (8a) and (8b) can easily be modified by an attribute. In that case the attribute + noun denotes to a set:

(10) Hungarian \hspace{1cm} English
a. Van piros alma? \hspace{1cm} a. Are there red apples?
COP.3SG red apple
‘Are there red apples?’
b. Vastag könyv-et olvas. \hspace{1cm} b. He is reading *( a) thick book.
thick book-ACC read.3SG.1f
‘(S)he is reading a thick book.’

Interestingly, the set-like nature of the noun is preserved if the nominal head is specified by a possessor. Consider for instance the use of könyvem [book-1SG.POSS] ‘my book’ in (11a) in contrast with (11b):

(11) a. Könyv-em az asztal-on van.
book-1SG.POSS the table-SUP.ES COP.3SG
‘There are books of mine on the table’
[lit. There is my book on the table]
b. A könyv-em az asztal-on van.
the book-1SG.POSS the table-SUP.ES COP.3SG
‘My book is on the table.’

Nouns marked by a possessive suffix behave as bare nouns in the sense that they combine with articles and numerals in the same way as bare nouns. They also allow for a collective marker, which, however, takes a different form, -i, instead of -k.

(12) a. egy könyv-em
a book1SG.POSS
‘a book of mine’
b. két könyv-em
two book-1SG.POSS
‘two books of mine’
c. könyv-ei-m
book-COLL-1SG.POSS
‘books of mine’

Another relevant property of forms such as (a) könyvem ‘my book’ as in (11) is that they are based on two Referential Acts. Schematically:

(13) a. (-id R₁: [(Tᵢ₁) (+id R₃: [+S, –A] (Rᵢ₃)) (Rᵢ₃)]) = könyvem (8a)

b. (+id R₁: [(Tᵢ₁) (+id R₃: [+S, –A] (Rᵢ₃)) (Rᵢ₃)]) = a könyvem (8b)

Note that both referents in könyvem count as identifiable (+id), as shown in (13b). Without the article, however, the referent is not identifiable (-id), whereas the possessive restrictor is (+id), as shown in (13a). The non-identifiability of the noun phrase is a condition for noun phrases to refer to sets and not to entities. The data discussed in this paragraph will be of particular interest in Section 5 below, where the status of the embedded possessor being (+id) will be further explored.

4.2. Articles
Articles in Hungarian are used to individuate or to refer to entities. Hungarian has both an indefinite and a definite article.⁵ The indefinite article egy is identical to the numeral ‘one’ (cf. 9a), the only difference between them being one of stress (unstressed for the article, and stressed for the numeral) and the simple (for the article) versus geminate (for the numeral) palatal stop (rendered as gy in the orthography): thus /egy/ ‘a’ versus /eggy / ‘one’ (cf. Kenesei et al. 1998: 95). The indefinite article is used to instruct the Addressee to construe a referent in conversational space, whereas the definite article is used if a referent has already been established and only needs to be retrieved.

(14) Hungarian                                      English
a. egy /egy/ könyv                       a. a book
   a book
   ‘a book’

b. a könyv                       b. the book
   the book
   ‘the book’

c. a könyv-ek
   the book-COLL
   ‘the books’

d. a két könyv
   the two book
   ‘the two books’

d. the two book-PL
   ‘the books’

The definite article combines with numerals (14d), but not with egy ‘one’. If number is one, Hungarian employs the nominal form egyik ‘one’ or the adjectival form egyetlen ‘one/single’ in combination with the definite article.⁶
5. ‘Non-identifiable objects’ and the use of the second form

So far the distribution of the two sets of person markers seems rather straightforward: 1f is used when the referent is not considered identifiable, 2f is used when it is. In this section, however, two cases are discussed in which the verb takes the second form even though the object does not obviously qualify for the status of (+id): embedded indefinite possessives and interrogative melyik ‘which’.

5.1. Embedded possessives

Section 4 presented data illustrating that Hungarian has set nouns, and that set nouns can have cardinality. Nouns can also be restricted by a possessor. Noun phrases containing a noun and a numeral, or a noun, a possessive restrictor and a numeral are not used to indicate that the referent is identifiable, as illustrated in the following examples. In neither (16a) nor (16b) does the speaker assume that the object of what ‘we’ read will be identifiable to the addressee. Note, however, that the first form is used in (16a) but the second form in (16b).

   two book-ACC read-PST-1PL.1f
   ‘We read two books.’

   b. Két könyv-é-t olvas-t-uk.
   two book-3SG.POSS-ACC read-PST-1PL.2f
   ‘We read two of his books.’

How can we explain the use of the second form in (16b)? Clearly it cannot be triggered by the object argument being identifiable, because it is not. One may wonder whether the second form is used because the embedded possessor is identifiable. This will be shown to be correct. In De Groot (1989) I argued that in cases such as (16b) the second form is not controlled by definiteness in the strict sense but by the related notion of ‘restricted superset’, based on the following consideration. The difference between the objects in (16a) and (16b) is that in (16b) the number of possible referents is limited, namely ‘the books somebody owns or has written’, whereas this restriction does not hold for the object of (16a), where all books in the world are potential referents. In the remainder of this section I will show that this view is not correct and that the identifiability of the embedded possessor expressed by a referential element on the noun is the crucial factor.

Before doing so, let me first provide some more information about possessive constructions such as the ones in (16). In Hungarian, the possessed is obligatorily combined with a referential suffix denoting the possessor. The possessor may also be lexically expressed as an apposition to the person marking element on the possessed. This is illustrated in the following pair of examples:

(17) a. két könyv-e
   two book-3SG.POSS
   ‘two books of him/her’

   b. János-nak két könyv-e
   John-DAT two book-3SG.POSS
   [of John two his book]
   ‘two books of John’
It needs to be pointed out that the possessive person marker is a referential marker and not an agreement marker (cf. De Groot 1989, ch. 3; De Groot & Limburg 1986). For expressions such as (17b) three Referential Acts must be postulated at the Interpersonal Level:

(18) \((-\text{id \ R}_I: [\langle \text{T}_I \rangle, (+\text{id \ R}_J: [-\text{S}, -\text{A}] (\text{R}_J)), (+\text{id \ R}_K: \text{János} (\text{R}_K))] (\text{R}_I))\)

könyv 'book'
-e 'his'
János 'John'

The coreferentiality within the possessive construction will be accounted for by co-indexation at the Representational Level. The Referential Act introducing the possessor, \(\text{R}_J\), will always be (+id): the possessor in the adnominal possessive construction is always presupposed and therefore identifiable, either because the speaker considers the possessor referent to be directly available to the addressee, or because the third Referential Act makes the possessor referent available. This third referent may be identifiable, as in (18), or not, as in (19).

(19) a. egy *holland-nak* két könyv-e
a Dutchman-DAT two book-3SG.POSS
[of a Dutchman two his book]
‘two books of a Dutchman’

b. (-id \(\text{R}_J: [\langle \text{T}_I \rangle (+\text{id \ R}_J: [-\text{S}, -\text{A}] (\text{R}_J)), (-\text{id \ R}_K: \text{holland}(\text{R}_K))] (\text{R}_I))\)

könyv 'book'
e 'his'
holland 'Dutchman'

Note that in (19) there is only one Referential Act with the specification (+id), which corresponds to the possessive person marker on the noun. If (19) formed the object, the verb would take the second form:

(20) Két könyv-ét olvas-t-uk egy holland-nak
two book-3SG.POSS-ACC read-PST-1PL.2f a Dutch-DAT
‘We read two books of a Dutchman.’

Before we can conclude that it is the identifiable possessor which relates to the use of the second form and not any other identifiable embedded restrictor, consider the following example, where the embedded restrictor in the object noun phrase contains identifiable material (but not a restrictor) and the second form is not used, but the first form:

(21) Két könyv-et olvas-t-ünk, amely a polc-on van.
two book-ACC read-PST-1PL.1f which the shelf-SUP.ES COP.3SG
‘We read two books, which stand on the shelf.’

Example (21) also illustrates that the notion of ‘restricted superset’ is not useful here, because (21) uses the second form, despite the fact that the notion of restricted superset (i.e. the limited number of books which fits on the bookshelf referred to) applies.

It is thus safe to conclude that the second form is used in Hungarian if the object phrase contains a possessor, because possessors, expressed by referential suffixes, are always identifiable. The representation in Functional Discourse Grammar of a non-identifiable noun phrase which contains an identifiable possessor which will trigger the use of the second form could be as follows (Hengeveld & Mackenzie 2008: 243):^7
(22) Két könyv-é-t (olvastuk).
   two book-3SG.POSS-ACC (read.PST.1PL.2f)
   ‘(We read) two of his/her books’

   IL: (-id R₁; [(T₁), (+id R₂; [–S, –A] (R₃))] (R₄))
   RL: (2 xᵢ: [(fᵢ: könyv[N] (fᵢ))] (xᵢ): [(xᵢ)_Ass] (xᵢ))
   ML: second form

5.2. Interrogative melyik ‘which’
Interrogative words are often considered indefinite pronouns in Hungarian. This is why grammars of Hungarian generally use this criterion to explain the application of the first form when the object contains an interrogative word.\(^8\) I agree that noun phrases with an interrogative word cannot be used as noun phrases to indicate that the referent is identifiable (cf. (15)). This general claim, however, does not hold for the use of melyik ‘which’, as can be witnessed from the following example, where the second form applies:

(23) Melyik almá-t akar-od?
   which apple-ACC want-2SG.2f
   ‘Which apple do you want?’

   Melyik ‘which’ invites the addressee to make explicit which apple of a set of different apples (identifiable to speaker and addressee) s/he wants by pointing at (e.g., ‘that one’) or describing it (e.g., ‘the red one’).\(^9\) In this case the set is identifiable to speaker and addressee, whereas the entity to be chosen is not. Contrary to the claim I have made for embedded possessors, in this case I prefer the analysis in which the notion of restricted superset is relevant. I would therefore like to introduce the operator of ‘contextual identifiability’ [c+id] at the Interpersonal Level. This operator is relevant to noun phrases such as ‘which apple’, as in (23), where the set of apples is identifiable, but also, for instance, in cases of inference, as in It is a nice house, but the kitchen is too small (Rijkhoff 1989). The following specification of the three levels holds for example (23).\(^10\)

(24) IL: (c+id –s R₁: [(T₁)] (R₁))
    RL: (xᵢ: alma[N] (xᵢ))
    ML: second form

6. Identifiable objects and the use of the first form
In addition to these earlier counterexamples where apparent not-identifiable referents took 2f, there are also cases that force the use of 1f regardless of identifiability. This is the case with first and second person referents, which must always take 1f, despite the fact that Speaker and Addressee appear to be the prime candidates for identifiable referents in discourse. The polite form used to refer to a second person, however, is an exception to this rule. Object relative clauses, too, must take 1f cross-reference markers regardless of information status.

6.1. First and second person versus third person.
Hungarian makes a distinction between first and second person object on the one hand and third person object on the other. The former select the first form, whereas the latter takes 1f or 2f depending on its status in the discourse. Compare:
The dichotomy between first/second person versus third person cannot be accounted for in terms of identifiability. If identifiability were a decisive factor here, first and second person, speaker and addressee respectively, should be prime candidates for being identifiable. Instead, the notion of discourse participant seems to be relevant here. It could be argued that the person distinctions associated with the discourse participants, in particular speaker and addressee, are treated differently from person distinctions which are associated with non-discourse participants. Crucially, identifiability distinctions are not relevant to first and second person, whereas they are relevant to third person. There is no opposition possible between an identifiable ‘me’ or ‘you’ and a non-identifiable ‘me’ or ‘you’, whereas such opposition is possible in the case of third person, such as for instance ‘he’ (+id) versus ‘somebody’ (-id), which gets 1f. The model of Functional Discourse Grammar can deal with the different status of first and second person on the one hand and third person on the other in an insightful manner. The reasoning is as follows. Consider the general structure of a Discourse Act, where the restrictor of the A-variable consists of an illocutionary frame:

\[
(\Pi A_1; [(\Pi F_1; ILL (F_1)) (P_1)_S (P_2)_A (C_1))] (A_1))
\]

The illocutionary frame consists of an Illocution F, which is presented as an abstract three place predicate with its arguments, (P_1)_S, (P_1)_A and (C_1), i.e. Speaker, Addressee, and Communicated Content. Within the Act the Communicated Content contains everything the Speaker wishes to evoke in his/her communication by means of using Subacts to evoke a property (Subact of Ascription) or a referent (Subact of Reference); schematically this is represented as follows:

\[
(\Pi C_1; [ \ldots (\Pi T_1)_\Phi (\Pi R_1)_\Phi \ldots ] (C_1))_\Phi
\]

The model assumes that the referents of first and second person pronouns in the Communicated Content are identified through co-reference with the two participants in the Illocution, whereas third person pronouns lack such type of co-reference. Consequently, first and second person cannot be further specified by +/-id operators, whereas third person can. These factors neatly account for the possibility in Hungarian to treat first and second person differently from third person. First and second person can be associated with arguments of the Illocutionary predicate, whereas third person can only be associated with Referential Subacts:

\[
(\Pi A_1; [(\Pi F_1; ILL (F_1)) (P_1)_S (P_2)_A (\Pi C_1; [(\Pi T_1)_\Phi (\Pi R_1)_\Phi [C_1])]_\Phi) (A_1))
\]

Of course, first and second person can be introduced as Referential Subacts. Still, they are associated with the P_1 and P_2 by means of co-referentiality. Third person, which cannot be associated with on of the discourse participants, allows different identifiability distinctions. Rather than the formalizations presented in the introductory chapter by Keizer & Van Staden (this volume), I would like to suggest the following representations of person on the Interpersonal Level:
first person singular: \((R_1: [+S, –A] (R_1))\) ‘I’
second person singular: \((R_1: [–S, +A] (R_1))\) ‘you’
third person singular: \((+id R_1: [–S, –A] (R_1))\) ‘he’
third person singular: \((-id R_1: [–S, –A] (R_1))\) ‘somebody’

The fact that the status of the first and second person is insensitive to identifiability distinctions forms the basis of an account for the application of the first form in Hungarian with objects specified for first or second person distinctions. The first form could be considered the default, which applies when there is no object, the object is not specified, the object is not specified for identifiability, or the object counts as non-identifiable. The second form is triggered by just one distinctive feature: the object counts as identifiable.

There is, however, an exception to this rule. When the addressee is referred to by means of the polite forms maga/maguk and ön/önök ‘you.SG/you.PL’, the verb takes 2f rather than expected 1f. Compare the following example with (25) above.

\(30\) Lát-ja magá-t / ön-t // maguk-at / önök-et.
See-3SG.2f you(polite).SG-ACC // you(polite).PL-ACC
‘(S)he sees you (polite singular) / you (polite plural).’

The explanation for this lies in the history of the forms. The form maga – first attested in 1195 – originates from a noun + third person singular possessive marker (mag-a body-3SG.POSS ‘his body’). From this meaning developed the reflexive ‘himself’ and later the polite form (Benkő 1976 part 2: 811). The reflexive in present-day Hungarian still takes the form of maga where the final –a counts as the third person possessive marker.\(^{12}\) The polite personal pronouns maga and maguk have the same form as the third person reflexive pronouns singular and plural, respectively. As argued in Section 5.1 above, the presence of a possessive as an embedded identifiable referent triggers the second form. This consequently also holds for the polite personal pronouns, because they contain a possessive referential element. The polite forms ön/önök, however, have been artificially introduced into grammar in 1806, during the period of the language reform. At that time Hungarian language reformers often took German as a model after which the Hungarian language could be moulded to become a ‘full-fledged modern European language’. One such novelty was a unique polite form (in addition to the reflexives which also functioned as polite forms), which in German is treated as a third person plural.\(^{13}\) In the world’s languages, polite forms often take a form which expresses more distance between speaker and addressee. Using the third person form to address somebody creates this distance. One could argue for different treatments of such polite forms in Functional Discourse Grammar. The first one is based on semantic considerations and treats the politely addressed Addressee as a third person:

\(31\) third person singular [polite]: \((+id R_1: [–S, –A] (R_1))\)

The second one is based on morphosyntactic considerations; in this case, the combination of the politeness operator [h] with an Addressee analysed as a second person will lead to a third person format the Morphosyntactic Level.

\(32\) second person singular [polite]: \((hR_1: [–S, +A] (R_1))\)

In the case of Hungarian I opt for the latter approach, for the following reasons. The polite forms ön/önök are clearly a newly created, which rather calls for a treatment at the
Morphosyntactic Level and not at the Interpersonal Level. Given their internal morphological structure this also holds for maga/maguk. The use of the second form with polite objects would in this case then be an instance of a grammaticalized form and not the expression of identifiability based on a choice made at the Interpersonal Level.

6.2. Relativized objects

A second example of a mismatch between an identifiable object and the first form is found in relative clauses. The finite verb of a clause relativized from the object takes the first form and not the second one:

(33) a film, ami-t Mari tegnap láttott
the film that-ACC Mary yesterday see-PAST.3SG.1f
‘the film that Mary saw yesterday’

The referent denoted by the relative pronoun must be taken to be identifiable since it refers to the antecedent, the film. The history of Hungarian, however, reveals that relative pronouns are derived from interrogative pronouns. The origins of the interrogative pronouns go back to the very early stages of Hungarian. They count as indefinites in the sense that in the function of object they combine with the first form of the verb and not the second form. In later stages, the interrogative pronouns were combined with elements such as vala- (probably a form of the existential verb ‘be’) into indefinite pronouns, as valaki ‘somebody’, valami ‘something’, and valahol ‘somewhere’. Only relatively recently, i.e., in the fifteenth/sixteenth century, the interrogative pronouns were also employed as relative pronouns. Consider the following example from 1526:

(34) ‘Old Hungarian’ (Benkő 1992: 228)
az mith yrtal azt megh ertetem
that what.ACC write.PST.2SG.1f that.ACC ASP understand.PST.1SG.2f
‘I understood what you wrote.’

Later the demonstrative az ‘that’ and the indefinite pronouns fused together in single words in the following way (Bárczi et al. 1978: 210):

(35) a. az ‘that’ + ki ‘who’ -> azki -> akki -> aki ‘who (REL)’
   b. az ‘that’ + mi ‘what’ -> azmi -> ammi -> ami ‘that (REL)’
   c. az ‘that’ + hol ‘where’ -> azhol -> ahhol -> ahol ‘where (REL)’

Although the indefinite pronouns were combined with demonstratives, the verbal form did not change along with the pronouns. Moreover, relative pronouns could be used and still may be used without an antecedent, as indefinite relative pronouns, as for instance in:

(36) Aki korán kell, aranyat lel.
who early get up gold.ACC find
[lit. Who gets up early will find gold]
‘The early bird catches the worm.’
(37) a. Mi-t  lát-sz?  ‘What do you see?’
   what-ACC see-2SG.1f
b. Valami-t  lát-sz.  ‘You see something.’
something-ACC see-2SG.1f
c. Ami-t  lát-sz, ....  ‘What you see, …’
   what.REL-ACC see-2SG.1f

I conclude that the use of the first form in cases where the object is specified by a relative pronoun is not based on a choice made at the Interpersonal Level, because the referent of the entity referred to by the relative pronoun is fully identifiable. The use of the first form is based on a grammatical property of relative pronouns, namely that they belong to the part of speech of indefinite pronouns.

Irrespective of the fact that, from a communicative point of view, relative pronouns are definite in the sense that the referent is identifiable, the part of speech remains that of indefinite pronoun. This grammatical aspect seems to carry more weight in determining the choice of verbal form than the pragmatic status of the relative pronoun.

(38) a.  film,  ami-t  lát-ott
   the   film,   REL-ACC see-PST.3SG.1f
   ‘the film that (s)he saw’

IL:  (+id R1: [(T1j) (Rj): [(Tj) (Rj): [-S, -A] (Rj))]) (Rj))
RL:  (1 xj: film[N] (xj): [PAST see[V] (xj)Exp (xj)U])
ML:  Np  →  [REL ACC]

First form

Note that at the Interpersonal Level there are two Referential Subacts: the NP (Rj) and the co-referential object argument within the relative clause (Rj). Also note that the co-referential (Rj) does not have the +id operator; hence the first form applies.

7. Object markers, definiteness, and pragmatic functions

The marking of identifiable objects by means of a kind of agreement on the verb is a rather rare phenomenon in the languages of the world. It is mainly found in the (related) Ugric, Samoyed and Mordvin languages. The marking of objects in the context of definiteness is, however, found more often. Examples of such marking devices are for instance the use of a (resumptive) pronoun (e.g. ‘clitic doubling’ in Greek), a special marker (e.g. Persian rā), and case marking (e.g. Estonian genitive ‘definite’ vs. partitive ‘indefinite’). These markers, however, should not necessarily be considered markers of definiteness of the object. The genitive vs. partitive case opposition of the object used with transitive verbs creates a telicity contrast in Estonian, just as in Finnish. Under a telic aspect the object generally counts as definite; under an atelic aspect as indefinite. The resumptive pronoun in Greek is not a definite marker in itself. It is used when there is a topical definite object preceding the verb. It resembles an earlier state of Hungarian as indicated in Stage 2 of example (6) above. The object marker rā in Persian is often considered a definite marker. This view has recently been challenged in favour of rā being a topic marker or a marker signalling that the definite NP is not the subject of the clause (cf. Van Staden & Seyed-Gohrab 2004). Note that the pragmatic notion of topic or topicality often occurs within the discussion of phenomena related to definite objects, as in Greek and Persian. When we consider Hungarian, there is no reason to assume that a relation holds between identifiable objects and topicality. Topical elements are
placed in clause initial position. Identifiable objects may appear in that position, but may also
turn up elsewhere in the clause, even in focus position (cf. 39b). Compare in this respect the
following three sentences:

(39) a. Péter látta a fá-t.
Peter saw.PST.3SG.2f the tree-ACC
‘Peter saw the tree.’

b. Péter A FÁ-T látta.
Peter the tree-ACC saw.PST.3SG.2f
‘It was the tree that Peter saw.’

c. A fá-t PÉTER látta.
The tree-ACC Peter saw.PST.3SG.2f
‘The tree, Peter saw it.’

8. Formalization

The facts described above show that a direct relation could be established between person
distinctions and identifiability at the Interpersonal Level and the use of the two sets of cross-
referencing markers in Hungarian. This relation holds between first and second person objects
and the first form, and between (non)identifiable (third person) objects and the use of the first
or second form. These facts may be captured by the following formalizations in a FDG of
Hungarian.

The general structure of an Discourse Act is:

(40) (Π A₁: [(Π F₁: ILL (F₁)) (P₁₈) (P₂₈) (Π C₁: [(Π T₁) (Π R₁) (C₁)])] (C₁₈): (A₁)

By performing Referential Subacts, the Speaker may take him/herself, the Addressee or
somebody/thing else as the referent, leading to different specifications of (Π R₁)Φ:

(41) a. (R₁: [+S, –A]) (R₁)
‘first person’
b. (R₁: [–S, +A]) (R₁)
‘second person’
c. (hR₁: [–S, +A]) (R₁)
‘second person (polite)’
d. (±id R₁: [–S, –A]) (R₁)
‘third person’

In the case of non-pronominal forms, the Referential Subact will be specified by one or more
Ascriptive Subacts, as in (40):

(42) a. (−id R₁: [(T₁)] (R₁))
e.g. ‘a car’
b. (+id R₁: [(T₁)] (R₁))
e.g. ‘the car’
c. (+id R₁: [(T₁) (T₂)] (R₁))
e.g. ‘the red car’

The constellation of Referential Subacts specified for identifiability (i.e. operator (−id) or
(+id) at the Interpersonal Level) will trigger the first or second form of the finite verb at the
Morphosyntactic Level. Example (43) illustrates the first form and (44) the second form.

(43) Lát-ok egy kocsit.
See-1SG.1f a car.ACC
‘I see a car.’

IL: (C₁: [(T₁) (R₁: [+S, –A]) (R₁)] (−id R₁: [(T₁)] (R₁)) (C₁))
ML: first form
(44) Lát-om  a kocsit.
See-1SG.2f the car.ACC
‘I see the car.’

IL: \( (C_1: [(T_1) (R_1: [+S, –A]) (R_3)]) (+id R_1: [(T_3)] (R_3)) (C_0) \)
ML: second form

First and second person with the function of Undergoer do not trigger the second form because they are not marked for identifiability at the Interpersonal Level.
The polite form can now be represented as in (45):

(45) Lát-om Önt.
See-1SG.2f you(polite).ACC
‘I see you.’

IL: \( (C_1: [(T_1) (R_1: [+S, –A]) (R_3)]) (hR_1: [–S, +A]) (R_3) (C_0) \)
ML: second form

9. Conclusions
This paper has offered a detailed description of the use of verbal cross-referencing markers in Hungarian. On the basis of the discussion of the data against the background of a linguistic theory, viz. Functional Discourse Grammar, the following conclusions can be drawn.
(i) The second form in Hungarian can indeed be taken to be the expression of identifiability of the object at the Interpersonal Level. The first form could be considered a default, i.e. the form used in all other cases.
(ii) Four apparent counterexamples are accounted for in the following fashion. In the case of the relative pronoun as an indefinite pronoun, and the polite form as a third person, grammaticalization has been show to play a role. First and second person do not allow for identifiability distinctions, they are always identifiable by way of co-indexation with Speaker and Addressee in the illocutionary frame. Therefore the default form applies if the object is specified for first or second person. In the case of _melyik_ ‘which’ it is argued that the context is identifiable for which the operator of ‘contextual identifiability’ is introduced. Finally, if the object noun phrase contains a possessive restrictor, the second form applies because the possessor is identifiable.
(iii) Not only bare nouns may have the property of the _Seinsart_ ‘set noun’ but some larger syntactic units as well. Nouns specified by an attribute or a possessor may still have the properties of ‘set noun’.
(iv) Identifiability applies both to specific entities as to sets. They may both trigger the second form on the finite verb.
(v) First and second person distinctions in FDG are not sensitive to identifiability operators and should be kept apart from third person distinctions which are.
(vi) Inconsistencies in the use of the first and second form of the verb in Hungarian can be explained in terms of grammaticalization. Other aspects in grammar prevail over the interpersonal aspect of identifiability. This is the case of the use of the polite form (one based on the third person reflexive form, and the other on German) and the relative pronoun, which is considered an indefinite pronoun.
Appendix 1. Verbal paradigm

All forms of öl ‘kill’. The lexical entry of verbs in Hungarian is the form of third person singular, present tense, indicative. This form is actually the stem of the verb.

‘First form’ refers to the set of cross-referencing markers that is used with intransitives and transitives when the object counts as non-identifiable, and ‘second form’ to the set of cross-referencing markers that is used with transitives when the object counts as identifiable. The forms that do not fall under one of these headings are the forms that are used when the subject is first person singular and the object second person singular or plural (cf. Section 3, example (7)).

<table>
<thead>
<tr>
<th>INDICATIVE</th>
<th>Present</th>
<th>Past</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>first form</td>
<td>second form</td>
<td>first form</td>
</tr>
<tr>
<td>ölok</td>
<td>ölm</td>
<td>ötem</td>
<td>ötem</td>
</tr>
<tr>
<td>ölsz</td>
<td>ölod</td>
<td>ötél</td>
<td>öted</td>
</tr>
<tr>
<td>öl</td>
<td>öli</td>
<td>öte</td>
<td>ötek</td>
</tr>
<tr>
<td>öltunk</td>
<td>öltük</td>
<td>öltünk</td>
<td>öltük</td>
</tr>
<tr>
<td>öltök</td>
<td>ölték</td>
<td>öltetek</td>
<td>öltétek</td>
</tr>
<tr>
<td>ölnek</td>
<td>ölik</td>
<td>ötek</td>
<td>ötek</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUBJUNCTIVE</th>
<th>Present</th>
<th>Past</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>first form</td>
<td>second form</td>
<td>first form</td>
</tr>
<tr>
<td>öljelek</td>
<td>öljem</td>
<td>öltelek</td>
<td>ölni foglak</td>
</tr>
<tr>
<td>öljél</td>
<td>öljed</td>
<td>öltelek</td>
<td>ölni foglak</td>
</tr>
<tr>
<td>öljön</td>
<td>öje</td>
<td>öltelek</td>
<td>ölni foglak</td>
</tr>
<tr>
<td>öljünk</td>
<td>öljük</td>
<td>öltelek</td>
<td>ölni foglak</td>
</tr>
<tr>
<td>öljétek</td>
<td>öljétek</td>
<td>öltelek</td>
<td>ölni foglak</td>
</tr>
<tr>
<td>öljenek</td>
<td>öljék</td>
<td>öltelek</td>
<td>ölni foglak</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CONDITIONAL</th>
<th>Present</th>
<th>Past</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>first form</td>
<td>second form</td>
<td>first form</td>
</tr>
<tr>
<td>ölnék</td>
<td>ölném</td>
<td>ötem volna</td>
<td>ötem volna</td>
</tr>
<tr>
<td>önlél</td>
<td>önléd</td>
<td>ötlél volna</td>
<td>öted volna</td>
</tr>
<tr>
<td>önléne</td>
<td>önlé</td>
<td>ölt volna</td>
<td>ölte volna</td>
</tr>
<tr>
<td>ölnénk</td>
<td>ölnénk</td>
<td>öltünk volna</td>
<td>öltük volna</td>
</tr>
<tr>
<td>ölnétek</td>
<td>ölnétek</td>
<td>öltetek</td>
<td>öltétek volna</td>
</tr>
<tr>
<td>ölnéneék</td>
<td>ölnéék</td>
<td>volna</td>
<td>ölték volna</td>
</tr>
<tr>
<td></td>
<td>önlélek</td>
<td>öltelek volna</td>
<td>ölni foglak</td>
</tr>
</tbody>
</table>

References


---

1 Abbreviations of theoretical terms and symbols used in this article are: IL (Interpersonal Level); RL (Representational Level); ML (Morphosyntactic Level); id (identifiability); c+id (contextual identifiability); R (Subact of Reference); T (Subact of Ascription); P (participant); h (honorific); S (Speaker); A (Addressee); A (Discourse Act); F (Illocution); C (Communicated Content); p (Proposition); e (State of Affairs); f (Property or Relation); x (Individual); m (plural); U (Undergoer); Exp (Experiencer); Ass (Associative).

2 Abbreviations of descriptive terms are: 1 (first person); 1f (first form); 2 (second person); 2f (second form); 3 (third person); ACC (Accusative); ASP (aspect); COLL (collective); COP (copula); DAT (dative); FUT (future); PL (plural); POSS (possessive); PST (past); REL (relative); SG (singular); SUP.ES (superessive).

3 Proper names can under certain pragmatic conditions be combined with the definite article in Hungarian. A discussion of these conditions, however, is beyond the scope of this paper.

4 As a matter of fact all Ugric, and also the Samoyed and Mordvin languages, distinguish between two conjugations depending on the absence or presence of an (in)definite object. Hajdú (1972: 44) firmly asserts that the distinction may be postulated for proto-Uralic.

5 See Givón (1976) and de Groot & Limburg (1986) for the general development of pronouns into clitics and agreement markers, and Hajdú (1972) for the historical development of Hungarian.

6 Interestingly, all set noun type languages in the sample of Rijkhoff (2002) have an indefinite and a definite article. In all cases the indefinite article is identical (or very similar) to the numeral ‘one’.

7 Evidence that the form egyik ‘one’ is indeed nominal is that the form takes nominal possessive endings, as in egyik-ük (one-3PL.POSS) ‘one of them’. Moreover, the form can be used independently with the definite article, as in az egyik (the one) ‘the one’. As for egyetlen, apart from the adjective, there is also a noun with the same form, meaning ‘the only one’. This form allows for all types of nominal flections and the use of the definite
article. In example (12), however, it is not the nominal form that is used but the adjectival form. It modifies the
nominal head of the phrase.

7 The semantic role of possessor is called ‘associative’ in Hengeveld & Mackenzie 2008 when the relation is
alienable, as it is here. Inalienable possessors are considered ‘referents’.

8 *Melyik* ‘which’ takes the position otherwise taken by ‘demonstratives’ in noun phrases (de Groot 2005a: 120).
The form requests a demonstrative in the answer. Kenesei et al. (1998) refer to question words such as *melyik*
‘which’ as selective interrogative pronouns opposed to general interrogative pronouns such as *ki* ‘who’ and *mi*
‘what’.

9 According to Hengeveld & Mackenzie (2008: 122-123), questioned elements are marked (+id, -s) at IL; the
referent is assumed identifiable for the Addressee but not for the Speaker.

10 I use the operators ‘-id’ and ‘+id’ to distinguish between third person identifiable and third person non-
identifiable in (29). I do not claim that these two operators are sufficient to account for distinctions made in other
languages. The third person pronouns of Navaho, for instance, are reported to differ in terms of definiteness,
with the less definite used in polite conversation (Young & Morgan 1964: 3). See also Siewierska (2004: 156).

11 Miklós Révai used the form *Ön* for the first time in his *Elaboratior Grammatica Hngarica* I, 302, 1806
(Benkő 1976, part 3: 27). The Hungarian grammarian Miklós Révai introduced the polite form in both singular
and plural on the basis of the bound morpheme *ön*–‘self’ as in *önmaga* ‘self’ or *önként* ‘voluntarily’, ‘of one’s
own free will’, as a free morpheme *ön* for the singular polite form and with the collective marker for the plural
*önnök*.

12 This development follows the general pattern found in the languages of Europe as observed by Heine &
Kuteva (2006: ch. 6).