Case and referential properties

Udo Klein[1] and Peter de Swart[2]

[1]Fakultät für Linguistik und Literaturwissenschaft
Universität Bielefeld
Universitätsstrasse 25
Postfach 10 01 31
33501 Bielefeld
Germany
Tel: +49 521 106-67102
Fax: +49 521 106-89050
udo.klein@uni-bielefeld.de

[2]Center for Language and Cognition Groningen
University of Groningen
PO BOX 716
9700 AS Groningen
The Netherlands
+31 (0)50 363-9114
P.J.F.de.Swart@rug.nl

Corresponding author: Udo Klein

Abstract

In this paper we discuss a number of languages with a multidimensional Differential Object Marking (DOM) system. In such languages overt object marking is determined by more than one argument feature such as animacy or definiteness. We will show that such argument features can be related to case marking in different ways. On the one hand, they can trigger the occurrence of overt marking, on the other they can be the result of it. We will demonstrate that different languages may prioritize the different argument features in different ways. These cross-linguistic patterns call for a more flexible approach to DOM than hitherto developed. We develop a sign-based declarative model that does not rely on hierarchies but instead accounts for language-specific patterns and cross-linguistic variation on the basis of feature structures.

Key words: differential object marking, hierarchies, animacy, definiteness, specificity

1 Introduction

In recent years the phenomenon of Differential Object Marking (DOM) has received considerable attention in the literature (Bossong, 1985; Aissen, 2003; Jäger, 2003; de Swart, 2007; von Heusinger, 2008; Malchukov, 2008, a.o.). In a language with a DOM-system direct objects are overtly (case) marked depending on certain features of the argument. Most commonly these features are the animacy, definiteness, or specificity of the object argument or any combination thereof. For instance, in Spanish the presence of the object marker *a* depends on the animacy and definiteness of the object and interacts with its specificity.

Most of the discussion concerning DOM has focused on the rationale underlying it. A recurrent analysis is one in terms of 'markedness reversal': what is unmarked for subjects is marked for objects and vice versa. Under this analysis objects that resemble prototypical subjects to the largest extent, i.e. animate and definite ones, are most likely to be marked. The naturalness or markedness of grammatical functions is often assessed in terms of ranking on a feature hierarchy. That is, the relevant features animacy and definiteness are reinterpreted as a hierarchy and features high on these hierarchies (e.g. human, definite) are typical for subjects whereas those low on the hierarchies (e.g. inanimate, indefinite) are typical for direct objects. With languages argued to choose different cut-off points, these hierarchies are not only used in the overall explanation but also in the language-particular descriptions.

Our goal in this article is not to explain the rationale behind DOM. Instead we will focus on the characterization of its manifestation in individual languages. The discussion will be mainly concerned with languages exhibiting a multidimensional DOM system in which more than one argument feature is involved. Our main aim is to show that although the same features play a role in different languages they do so in different ways. A recurrent theme will be that in specific DOM patterns those features that are inherent to a noun or a noun phrase take priority over those features that are not. We will point out what the implications of this observation are for theoretical accounts of DOM and we will develop a formal proposal for the interaction of factors in DOM systems.

In our view three different levels of investigation should be distinguished: (i) the characterization of language-specific patterns, (ii) the formulation of cross-linguistic generalizations, and (iii) the explanation of these cross-linguistic generalizations. It seems that in most recent work on DOM these three levels have been conflated into one. Contra to what seems to be generally assumed, we will claim that hier-

archies are not needed to characterize the DOM patterns found in the languages under discussion. In order to characterize those language-specific patterns we will make use of descriptive rules which themselves may be idiosyncratic in nature. We argue that hierarchies, if applicable at all, only enter at the higher level of language comparison, functioning as a type of *comparative concept* (Haspelmath, 2008a,b).

This article is organized as follows: in section 2 we discuss some data from languages with a multidimensional DOM system, and show that inherent and non-inherent features relate differently to case – inherent properties trigger object marking, while non-inherent properties are often the results of object marking. In section 3 we discuss some previous accounts of DOM, and in section 4 we present a rule-based analysis of multidimensional DOM within the formalism of sign grammar. Essentially, we postulate two rules for combining a direct object with a transitive verb, with optionality of object marking analysed as an overlap of the conditions under which rules apply. The formalism is flexible enough to accomodate the various types of split alternations, and accounts for fluid alternations by employing underspecification of features in the formulation of the rules. Finally, section 5 argues that the flexibility of the formalism is in fact necessary in order to characterize rare DOM patterns which provide exceptions to the cross-linguistic generalization about DOM, and that consequently cross-linguistic tendencies and generalizations should be understood as abstractions from the similar but not identical language-specific patterns. Section 6 concludes the article.

2 Multidimensional DOM

In this section we will discuss data from languages with a multidimensional DOM system. Our goal is to show that when in a language more than one factor interacts with overt object marking these factors may stand in a different relation to the object marker. A clear distinction should be made between factors that *trigger* the occurrence of overt object marking and the ones that are the *result* of the occurrence of overt object marking (cf. de Swart 2007; de Swart and de Hoop 2007). Triggers are properties that are either semantically or morphosyntactically intrinsic to an argument and are inert to change. These properties belong either to the head noun or to other lexical elements in the noun phrase (e.g. determiners); in both cases properties of individual items extend to the noun phrase as a whole. For instance, every noun has a given animacy value which we cannot change by adding or removing overt case marking from the argument. In other words, animacy is a property which is semantically inherent to a noun.

The animacy value of the head noun is inherited by the noun phrase, i.e. both the noun 'man' and the NP 'the man' have the feature of being human. Likewise, moving beyond the level of the noun to that of the noun phrase (or DP), DP-type or 'syntactic definiteness' may be a morphosyntactic argument feature that triggers case marking. In this case, the noun phrase inherits from one of its components (the determiner) a definiteness value. Hence, due to the presence of the definite article 'the' the NP 'the man' should be considered definite and it is this feature that triggers case marking. It is important to note that syntactic definiteness is used to refer to (the DP-type of) an argument and not to its semantic value. Thus, Danon (2001, 2006) has shown that in modern Hebrew it is not semantic definiteness that triggers the occurrence of the object marker *'et*, but rather whether it is syntactically marked as being definite. In the domain of nouns this (roughly) means that only those objects containing the definite article *ha* can be preceded by the object marker. Like animacy, syntactic definiteness or DP-type should thus be treated as a morphosyntactically intrinsic property of an argument, i.e. by adding or removing case from an argument we do not change its DP-type.

In addition to argument features that function as a trigger for overt case marking, we also find features that are the result of the use of overt case marking. Such 'result' features are properties that are non-inherent to an argument and are subject to change. This means we are dealing with features that are either not semantically intrinsic features of a noun or not morphosyntactically coded in the noun phrase. In other words, it is information that cannot be read off an argument (in isolation), even though it may be inferred on the basis of contextual information. In such situations case marking can be used to overtly code a feature on the NP. A recurring example of such a feature is the referentiality or specificity of an argument. It is well-known that in many languages the occurrence of overt case marking goes hand-in-hand with the specificity of an object. For instance, in Turkish by adding or removing case from an argument we can change its specificity (von Heusinger and Kornfilt, 2005).

What features function as triggers and which ones as results is partially language-specific and depends crucially on the morphosyntactic inventory of a given language. Animacy is a feature that qualifies as a universal trigger; in every language we can divide the set of nouns according to their animacy and we do not need overt marking to indicate the animacy value of a noun.[1] Definiteness, by contrast, only

[1]There is a limited set of nouns for which the animacy value in a given context may be the result of case marking. In the case of homonymous (or polysemous) nouns, overt case marking may distinguish between two different (animacy) readings, e.g. a lexical item like

functions as a trigger in those languages that have lexical means (e.g. determiners) to express it. If such devices are absent, definiteness may become a result feature, i.e. the consequence of the use of overt case marking. The same holds for specificity (although there may be less languages with determiners devoted to the encoding of the specificity of arguments).

This difference between triggers and results of object marking is closely related to the distinction between split and fluid case alternations proposed by de Hoop and Malchukov (2007). The so-called triggers are always involved in a split case alternation in which overt case marking literally makes a split between categories of a certain dimension. For instance, in a given language objects that are animate may be obligatorily marked whereas those that are inanimate are not. In this case absence of case marking on animate objects will result in ungrammaticality. The so-called result features, on the other hand, are always involved in a fluid case alternation in which the use of case applies within a category and has an effect on a dimension different from that category. For instance, accusative case may be used on inanimates in which case they are interpreted as specific or it may be absent in which case there is no claim with respect to their specificity. In a fluid case alternation the presence or absence of case does not correlate with (un)grammaticality but rather with a change in interpretation.

The contrast between triggers and results (and hence split and fluid alternations) may be better understood if we make a distinction between the speaker and hearer perspective. Features that function as a trigger force the speaker to use overt case marking, i.e. they play a role in language production. Result features, by contrast, emerge on the side of the hearer and hence are part of the interpretation process. Our terminology will become more apparent in Section 4 where we develop our formal analysis. There, 'triggers' will be shown to trigger the application of a certain grammatical rule, whereas 'results' will be shown to result from the application of such a rule.

In the remainder of this section we want to demonstrate that in multidimensional DOM systems inherent properties ('triggers') take priority over non-inherent ones ('results') and as a result that split alternations take priority over fluid ones. This means that fluid alternations can only occur in those areas of the grammar where split alternations leave room for them. Languages can of course display fluid marking without any split marking, but if a split marking pat-

'star' may be interpreted as referring to a human entity (a pop star) or a heavenly body depending on the absence or presence of case. We thank Klaus von Heusinger for this observation.

tern emerges in a language which already has a fluid marking pattern, there is a clear and obvious sense in which the split marking ends up dominating the fluid one. Consider a hypothetical language in which differential object marking on an NP implies that the argument is totally affected, and lack of case marking is compatible with the argument being totally affected, partially affected or not affected at all. Since all NPs are treated the same, this is not an instance of split but fluid marking. If the differential object marker gradually becomes obligatory on both affected and unaffected animate NPs, then despite the fluid origin of the marking the result is a split in the case marking of NPs: animates are marked obligatorily, whereas inanimates can but need not be marked, with a concommitant semantic implication that the argument is totally affected if marked. This is the sense in which the split marking can always be said to dominate fluid marking. Below we will first show that split alternations indeed take priority over fluid ones. Then we will demonstrate that when different triggers are involved they may prioritize differently in different languages. Finally, we will show that fluid alternations may not only be dominated by referential properties of arguments but also by more general grammatical principles.

It should be noted that in the discussion we will only focus on the indexing use of DOM and that we disregard its uses as an actual disambiguation mechanism. It is well-known that in most languages in which animacy is involved in DOM it is not only in an indexing way but also in a disambiguation way (de Swart, 2007; Malchukov, 2008; Primus, 2009). Thus, although we want to show that multidimensional DOM systems are more complex and varied than hitherto acknowledged, the reader should bear in mind that in reality the systems under discussion may be even more complex (see also Section 4).

2.1 Split over Fluid: Object Marking in Hindi and Kannada

In this section we will focus on the two-dimensional DOM systems of two South Asian languages, Hindi and Kannada. We start with Hindi, direct objects can be marked with *ko*, the same marker that is used for indirect objects. In the present discussion we limit ourselves to the use of *ko* on direct objects that occur without a determiner. The differential use of *ko* on direct objects has received much attention in the literature (see Mohanan 1990; Butt 1993; Singh 1994; McGregor 1995; Aissen 2003; de Hoop and Narasimhan 2005; Kachru 2006, a.o.) and two factors can be distinguished that influence it. On the one

hand, there is animacy as *ko* is obligatory for objects that are human, but not for objects that are animate or inanimate. On the other hand, the occurrence of *ko* is related to the definiteness or specificity of the direct object. Regarding the latter factor authors differ as to whether they take definiteness or specificity to be the primary factor. Mohanan (1990), for instance, seems to relate DOM in Hindi mainly to definiteness, with specificity playing a secondary role. Butt (1993), on the other hand, takes specificity to be the relevant notion but acknowledges that it interacts with definiteness. We will not make a principled choice for one or the other factor. Whether we call the interpretation given to a *ko*-marked direct object definite or specific, and that of an unmarked direct object indefinite or non-specific, does not affect our claim that animacy takes priority over definiteness/specificity in the use of *ko*.

Following Mohanan (1990), human objects have to be obligatorily marked with *ko*. When a human object is marked, it can be interpreted as definite or indefinite. When such an object occurs without *ko* this results in an ungrammatical sentence. This contrast is shown in (1) and (2) for the noun 'child':[2]

HINDI (Indo-Aryan; Mohanan 1990:103)
(1) Ilaa-ne bacce-ko uthayaa.
Ila-ERG child-KO lift.PF
'Ila lifted the/a child.'

(2) *Ilaa-ne baccaa uthayaa.
Ila-ERG child lift.PF

In the absence of a determiner, inanimate nouns, on the other hand, can either be marked with *ko* or be left unmarked. The use of *ko* does have repercussions for the interpretation associated with the direct object. An unmarked inanimate can be interpreted as definite or indefinite, as is shown in (3):

HINDI (Indo-Aryan; Mohanan 1990:103)
(3) Ilaa-ne haar uthaayaa.
Ila-ERG necklace lift.PF
'Ila lifted a/the necklace.'

[2]Abbreviations used: 1:first person, 3:third person, ABL:ablative, ACC:accusative, AGR:agreement, AOR:aorist, DAT:dative, DEF:definite, ERG:ergative, FEM:feminine, FUT:future, GEN:genitive, INF:infinitive, LOC:locative, MASC:masculine, NMZ:nominalizer, NOM:nominative, NPAST:non past, PF:perfective, PL:plural, PST:past, SG:singular.

Definiteness of an inanimate noun is expressed by using *ko*. This is shown for the noun 'necklace' in (4) below:

HINDI (Indo-Aryan; Mohanan 1990:104)
(4) Ilaa-ne haar-ko uthayaa.
Ila-ERG necklace-KO lift.PF
'Ila lifted the necklace.'

The above examples show that both animacy and definiteness play a role in differential object marking in Hindi. Their roles are, nevertheless, clearly differentiated. Consider the table in (5):

(5)

	human	-human
KO	def/indef	def
∅	*	def/indef

From this table we can conclude two things: (i) the use of *ko* on direct objects is primarily triggered by the humanness of the direct object. Absence of this marker on human direct objects results in ungrammaticality indicating that this is a split case alternation; (ii) definiteness does not trigger the use of *ko* but rather is an effect of the use of this marker, which means we are dealing with a fluid alternation. If we were to claim that definiteness triggers case marking on Hindi direct objects we would have trouble explaining why indefinite human objects are marked with *ko* as well. Furthermore, it is left unexplained why in the absence of case marking both a definite and an indefinite reading are possible for non-human objects. If definiteness triggers case marking we would expect a definite reading always to co-occur with *ko*.

A similar situation can be found in Kannada, a Dravidian language with a differential object marking system very similar to that of Hindi (cf. Lidz 1999, 2006). As in Hindi the occurrence of accusative case on direct objects interacts with the animacy and referentiality of the object. In Kannada, human and animate direct objects are obligatorily marked with accusative case. This is shown for human objects by the contrast in grammaticality between (6) and (7):[3]

KANNADA (Dravidian; Lidz 2006:11)
(6) *Naanu sekretari huDuk-utt-idd-eene.
I.NOM secretary look.for-NPST-be-1SG
'I am looking for a secretary.'

[3]The occurrence of the glide *v* ([w]) or *y* ([j]) in the initial position of the accusative ending is determined by the preceding vowel.

(7) Naanu sekretari-yannu huDuk-utt-idd-eene.
I.NOM secretary-ACC look.for-NPST-be-1SG
'I am looking for a secretary.'

Inanimate objects, on the other hand, can occur with or without accusative case:

KANNADA (Dravidian; Lidz 2006:11)

(8) Naanu pustaka huDuk-utt-idd-eene.
I.NOM book look.for-NPST-be-1SG
'I am looking for a book.'

(9) Naanu pustaka-vannu huDuk-utt-idd-eene.
I.NOM book-ACC look.for-NPST-be-1SG
'I am looking for a book.'

As for the interpretation of the direct objects, Lidz notes that an animate direct object marked with accusative case can either be interpreted as non-specific or specific (*de dicto* or *de re* in the terminology used by Lidz). The same holds for inanimate objects without accusative case. Inanimate objects which occur with accusative case have to be interpreted as specific (*de re*). The pattern is summarized in the table in (10):

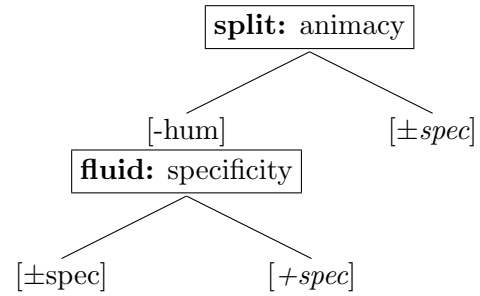
(10)

	animate	inanimate
ACC	*de dicto*/*de re*	*de re*
∅	*	*de dicto*/*de re*

This pattern looks very similar to that of Hindi, as again we find that an analysis of the accusative case as a specificity marker breaks down in the domain of animate direct objects. That is, it cannot be used as a specificity marker when it is required by the animacy of the direct object. In other words, animacy (a split alternation) takes priority over referentiality (a fluid alternation). As a result, the correlation between accusative case and a strong (definite/specific) interpretation does exist, but not across-the-board. It only holds in the domain of non-humans (Hindi) or inanimates (Kannada). This means that the fluid alternation can only apply where the split alternation leaves room for it.

The hierarchical relation between split and fluid alternations in Hindi and Kannada can be schematically depicted as in (11):

(11)



In this figure, the type of case alternation and the features involved are indicated in boxes. The hierarchical relation between alternations is reflected by dominance with the alternation taking priority dominating the other. Italicized terminal nodes indicate the application area of object marking. In (11) human objects which can be either specific or nonspecific are marked as are specific non-humans. In order to keep these figures as reader-friendly as possible terminal nodes are only specified for the most deeply embedded feature. It should be noted that they do however inherit the features of the nodes dominating them. Thus, the specification [+spec] of the node at the bottom right should be read as [-hum,+spec]. As will be discussed further in the next section, we assume that the patterns under consideration can be described through exclusive reference to binary features, which is reflected in the binary-branching tree structure used.

2.2 Split over Split: DOM in Spanish, Mongolian, and Romanian

In the previous subsection we discussed DOM systems in which a split alternation dominates a fluid one. In the present subsection we consider systems in which we find multiple split alternations and a fluid one interacting. The languages we will examine are Spanish, Mongolian, and Romanian.

The distribution of the Spanish object marker *a* has been extensively discussed in the literature but still is not entirely understood (for discussion, see Brugé and Brugger 1996; Torrego 1998; Delbecque 2002; von Heusinger and Kaiser 2003; Leonetti 2004; Bleam 2005, among many others). On our interpretation the occurrence of *a* follows an intricate pattern in which different split alternations and a fluid case alternation interact. The factors underlying the case splits are animacy and syntactic definiteness, and the one underlying the fluid case alternation is specificity.

The primary split is between animate and inanimate noun phrases in that only the former can take the object marker. This contrast

between animate and inanimate objects can be seen by comparing (12)-(13) to (14):

Spanish (Romance; Brugé and Brugger 1996:3)

(12) Esta mañana he visto *(a) Juan/la hermana de
this morning have.1sg seen a Juan/the sister of
María.
María
'This morning I saw Juan/María's sister.'

(13) Esta mañana he visto *(a) mi perro.
this morning have.1sg seen a my dog
'This morning I saw my dog.'

(14) Esta mañana he visto (*a) la nueva iglesia.
this morning have.1sg seen a the new church
'This morning I saw the new church.'

Examples (12) and (13) show that human and animate objects have to be marked with the prepositional object marker, something which is prohibited for the inanimate object in (14).

The obligatoriness of the object marker with human and animate objects only holds if they are not preceded by an indefinite article. Hence, the second split case alternation in Spanish is determined by syntactic definiteness. Indefinite human and animate objects can occur with or without the object marker:

Spanish (Romance; Hopper and Thompson 1980:256)

(15) Celia quiere mirar un bailarín.
Celia want.3sg watch.inf a ballet.dancer
'Celia wants to watch a (non-specific) ballet dancer.'

(16) Celia quiere mirar a un bailarín.
Celia want.3sg watch.inf a a ballet.dancer
'Celia wants to watch (specific) a ballet dancer.'

The absence and presence of *a* correlate with a change in meaning. An indefinite object preceded by the object marker can be interpreted as specific, something which is not possible for an unmarked indefinite object.

In order to establish that lexical definiteness and specificity each play a separate role in Spanish DOM, consider the following examples:

Spanish (Romance; Leonetti 2004:83, García García 2005:23)

(17) Está buscando a alquien.
be.3SG looking.for A someone
'(S)he is looking for someone.'

(18) No está buscando a nadie.
not be.3SG looking.for A anyone
'(S)he is not looking for anyone.'

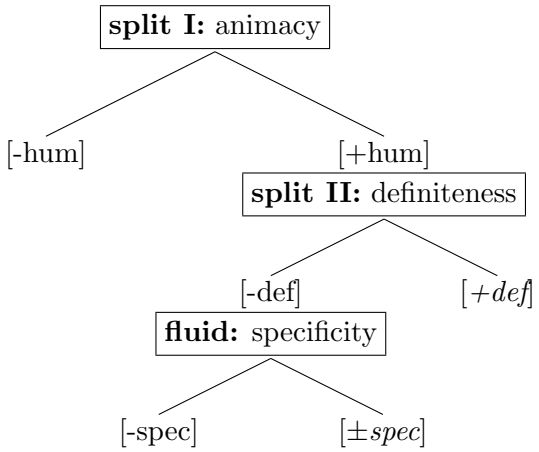
(19) Besó a todo el mundo.
kissed.3SG A whole the world
'(S)he kissed everybody.'

The examples in (17)-(19) all involve syntactically definite objects and have to be preceded by the object marker. Crucially, despite its presence none of the objects receive a specific interpretation. In fact, the examples all represent non-specific direct objects. This shows that syntactic definiteness is a factor independent from specificity. Furthermore, it shows that syntactic definiteness takes priority over specificity in determination of the occurrence of the object marker. Only when the direct object is not syntactically definite, can the object marker be used to indicate its specificity. Definiteness itself is in turn outweighed by animacy resulting in the following partial ordering of the relevance of the factors discussed so far: animacy > definiteness > specificity.

The correlation between specificity and the occurrence of *a* is in need of some further discussion. Leonetti (2004) argues that direct objects without *a* can only be interpreted as non-specific. Indefinite objects preceded by *a*, by contrast, can be interpreted as both specific and non-specific. If this is the case, the pattern in Spanish differs from the pattern found in Hindi and Kannada in which the absence of case goes with both a specific and a non-specific reading and the presence of case only with a specific reading.

Spanish thus presents a situation in which a split based on animacy takes priority over one based on definiteness. Moreover, specificity is involved in a fluid alternation which operates in the grammatical space left open by the split alternations. This is schematically depicted in (20):

(20)



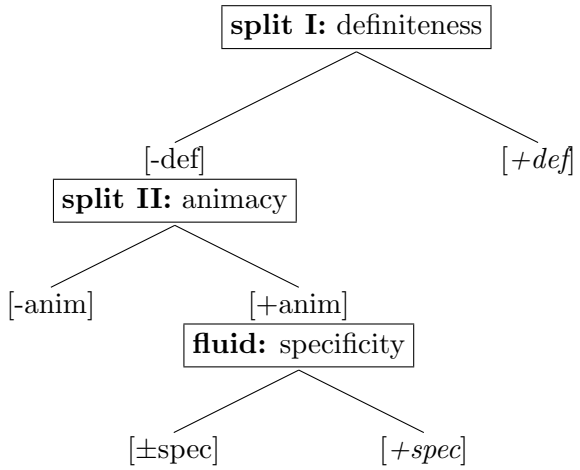
Mongolian presents a multidimensional DOM system in which we find the same type of split and fluid alternations as in Spanish (Guntsetseg, 2009). The splits are, however, ordered in a different way than in Spanish. In Mongolian all definite objects are obligatorily marked with accusative case irrespective of their animacy. Although it is generally assumed that Mongolian has no definite article, there are other elements (demonstratives, possessive affixes) that indicate definiteness (Guntsetseg, 2009). Accusative case is obligatory with objects marked with these elements. This means that the first split is based on definiteness. With indefinite noun phrases the use of the accusative case marker becomes dependent on animacy as it is only found with animate objects. This represents the second split. The object marker is however not obligatory with animate indefinite objects. Instead it is involved in a fluid alternation: when accusative case is used the animate indefinite has to be interpreted as specific, cf. (21), in absence of overt case marking it can be either specific or non-specific, cf. (22).

(21) Bold neg ohin-ig unssen.
Bold a girl-ACC kissed
'Bold kissed a certain girl.' [specific reading]

(22) Bold neg ohin unssen.
Bold a girl kissed
'Bold kissed a girl.' [specific or non-specific reading]

This means that the fluid alternation attested in Mongolian follows the general pattern we have seen several times by now. The full range of split and fluid alternations in this language can be schematically depicted as in (23).

(23)



Finally, we turn to Romanian, a language with an intricate interweaving of different case alternations. The first split is based on DP-type, more specifically pronominality, as we find the object marker *pe* as a rule with all pronouns, irrespective of the animacy of the referent:[4]

Romanian (Romance)

(24) Televiziunea m=a ales *(pe) mine, nu eu
television ACC.1SG=has chosen PE me not I
*(pe) ea.
PE 3SG.FEM
'Television has chosen me, not I it.'

Next, Romanian shows a complex interaction of different splits and fluid alternations. Simplifying somewhat (leaving out non-human animates which largely seem to follow humans except for indefinite NPs), DOM is prohibited with inanimate non-pronouns, cf. (27) and (28). For human objects it is obligatory for proper names and optional for both definite (25) and indefinite (26) NPs:

(25) (L=)am văzut (pe) copil-ul
ACC.MASC=have.1 seen PE child-DEF.MASC

[4]Romanian does not have a neuter (inanimate) pronoun like English *it*. Instead, pronominal gender matches the grammatical gender of the antecedent noun which is either masculine or feminine. In our analysis we rely on the referential animacy of a pronoun and hence include 'inanimate' as a category. If one prefers to stick to the grammatical animacy of pronouns, pronouns should be exclusively classified as human. In this case, the analysis could be simplified by leaving out the first split based on DP-type, making animacy the primary split. This analysis would result in a picture which is largely similar to the one sketched for Spanish above.

vecin-ului.
neighbor-DEF.MASC.GEN
'I/we have seen the neighbour's child.'

(26) (L=)am văzut (pe) un prieten.
ACC.MASC=have.1 seen PE a friend.MASC
'I/we have seen a friend.'

(27) (*L=)Am văzut (*pe) calculatorul-ul
ACC.MASC=have.1 seen PE computer-DEF.MASC
vecin-ului.
neighbor-DEF.MASC.GEN
'I/we have seen the neighbour's computer.'

(28) (*L=)Am văzut (*pe) un tractor.
ACC.MASC=have.1 seen PE a tractor.MASC
'I/we have seen a tractor.'

With indefinite human NPs Romanian displays a fluid alternation comparable to Hindi or Kannada. When the NP is object marked, it is interpreted as specific, and when it is not marked it can be interpreted either as specific or as non-specific.

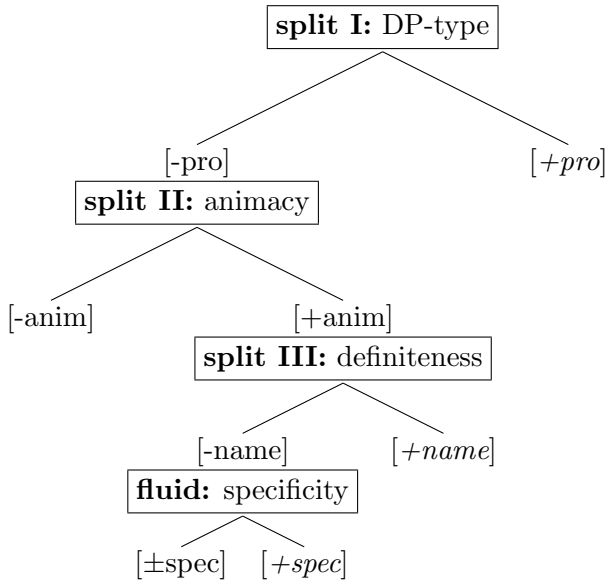
ROMANIAN (Romance; Dobrovie-Sorin 1994)
(29) Caut o secretară.
look-for.1.SG a.FEM secretary.FEM
'I'm looking for a (any or specific) secretary.'

(30) Caut pe o secretară.
look-for.1.SG PE a.FEM secretary.FEM
'I'm looking for a specific secretary.'

Definite human objects also seem to show a fluid alternation although it is not clear which semantic feature is being manipulated –suggestions include genericity (Dobrovie-Sorin 2007 as cited in von Heusinger and Onea 2008) and individualization (Stark, 2008)– making this an area for future research.

The DOM system of Romanian is schematically depicted in (31), abstracting away from the feature involved in the fluid alternation of definites.

(31)



2.3 Grammatical Principles over Fluid: Accusative Case in Turkish

In Turkish, accusative case on a direct object corresponds with a specific reading. The contrast between marked and unmarked direct objects is demonstrated in (32):

TURKISH (Turkic; von Heusinger and Kornfilt 2005:8)

(32) (Ben) bir kitap oku-du-m.
I a book read-PST-1SG
'I read a book.'

(33) (Ben) bir kitab-ı oku-du-m.
I a book-ACC read-PST-1SG
'I read a certain book.'

Von Heusinger and Kornfilt (2005) show, however, that accusative case is only a reliable indicator of specificity when the direct object immediately precedes the verb. In any other position, the use of accusative is obligatory and is compatible with a non-specific reading of the object. This is demonstrated in the following example in which the object 'tea' receives a non-specific (generic) reading:

TURKISH (Turkic; von Heusinger and Kornfilt 2005:11)

(34) Bizim ev-de çay-ı her.zamen aytül yap-ar.
our house-LOC tea-ACC always Aytül make-AOR
'Aytül always makes the tea in our family.'

A similar thing can be observed with the marking of embedded subjects. When they directly precede the verb and are unmarked they receive a non-specific reading, cf. (35), but when they are marked with genitive case they have to be interpreted as specific, cf. (36):

Turkish (Turkic; von Heusinger and Kornfilt 2005:15)

(35) ⌈Yol-dan bir araba geç-tiğ-in⌋-ı gör-dü-m.
road-ABL a car pass-NMZ-3SG-ACC see-PST-1SG
'I saw that a car [non-specific] went by on the road.'

(36) ⌈Yol-dan bir araba-nın geç-tiğ-in⌋-ı gör-dü-m.
road-ABL a car-GEN pass-NMZ-3SG-ACC see-PST-1SG
'I saw that a car [specific] went by on the road.'

Like direct objects, when the embedded subject is moved away from the preverbal position it has to be marked with genitive case and can receive either a specific or non-specific reading. This is illustrated in (37):

Turkish (Turkic; von Heusinger and Kornfilt 2005:16)

(37) ⌈bir araba*(-nın) yol-dan geç-tiğ-in⌋-ı gör-dü-m.
a car-GEN road-ABL pass-NMZ-3SG-ACC see-PST-1SG
'I saw that a car [non-specific or specific] went by on the road.'

Yet another environment in which the correlation between specificity and overt case marking breaks down involves partitives. When partitive direct objects occur with a lexical head they can surface with or without accusative case, resulting in the by now familiar difference in interpretation. This can be seen by comparing (38) and (39):

Turkish (Turkic; von Heusinger and Kornfilt 2005:32)

(38) Ali büro-ya çocuk-lar-dan iki kız al-acak.
Ali office-DAT child-PL-ABL two girl take-FUT
'Ali will hire, for the office, two (non-specific) girls of the children.'

(39) Ali büro-ya çocuk-lar-dan iki kız-ı al-acak.
Ali office-DAT child-PL-ABL two girl-ACC take-FUT
'Ali will hire, for the office, two (specific) girls of the children.'

When the lexical head of the partitive, *kız* in (38) and (39), is missing it has to be replaced by an agreement marker, *sin* in (40). This agreement marker, however, comes with the morphological requirement that it has to be followed by accusative case (in transitive contexts). Due to this formal requirement, accusative case can no longer

be used to indicate specificity and as a result partitive objects without a lexical head can be interpreted as both specific and non-specific. This is illustrated in (40):

TURKISH (Turkic; von Heusinger and Kornfilt 2005:34)

(40) Kitap-lar-dan iki-sin*(-ı) al, geri-sin-ı
book-PL-ABL two-AGR(3)-ACC buy remainder-AGR(3)-ACC
kutu-da bırak.
box-LOC leave
'Take (any) two of the books and leave the remainder [of the books] in the box.'

Thus, Turkish provides a good illustration of how formal requirements of the grammar, i.e., word order and agreement, can overrule the otherwise robust correlation between overt case marking and specificity. These formal requirements can be seen as split case alternations, e.g., a split between preverbal and non-preverbal position or between presence and absence of agreement morphology. In this way, the Turkish system provides additional evidence for the fact that split alternations take priority over fluid ones.

It is likely that these kinds of situations can be found in other languages as well. In Romanian, for instance, certain syntactic constructions like the comparative construction appear to require *pe*-marking irrespective of the animacy of the direct object (see von Heusinger and Onea 2008 for other constructions influencing the use of *pe*). Example (42) is the first hit found by Google for the search `site:ro "ca pe un"` on 13.07.2009. Like in Turkish the semantic import of a fluid alternation is lost in this case.

(41) Xcerion vede (*pe) OS-ul.
Xcerion sees OS-MASC.SG.DEF
'Xcerion sees the operating system.'

(42) Xcerion vede Web-ul ca pe un
Xcerion sees web-MASC.SG.DEF like PE a
OS.
operating.system
'Xcerion sees the web like (it sees) an operating system'.

(43) Xcerion vede WEb-ul ca un OS.
Xcerion sees web-MASC.SG.DEF like an operating.system
'Xcerion sees the web like an operating system (sees the web).'

The precise interaction of grammatical principles and case marking in DOM languages awaits further research.

3 Previous Accounts of DOM

In the previous section we have demonstrated that argument features can be related to overt object marking in different ways. Such features can be triggers that make the occurrence of overt case marking obligatory. Alternatively, certain argument features such as specificity can be the result of the occurrence of overt case marking. We have seen that the triggers are always involved in split case alternations, whereas the results partake in fluid alternations. Different languages may prioritize splits in a different way but they always let them dominate fluid alternations. The latter take up the grammatical space left over by split alternations. Thus, multidimensional DOM systems although perhaps looking similar to one another on the surface present a range of variation when one closely examines the interaction of different argument features. It is this variation that has to be captured in a formal approach to DOM.

In the literature on DOM we can distinguish two kinds of approaches. On the one hand, there are researchers who have mainly focused on the shifts in interpretation associated with the occurrence of overt object marking. In the terminology used here, they have concentrated on the fluid alternations in DOM systems. Indeed, many authors have proposed a systematic correlation between case and semantic interpretation (cf. Enç 1991; de Hoop 1992; Butt 1993; Ramchand 1997; Bleam 2005; Danon 2006). This correlation always seems to fall out in the following way: overt/accusative case corresponds with a strong interpretation, i.e., a definite, specific, *de re*, or presuppositional interpretation, and absence of case with a weak interpretation, i.e., an indefinite, non-specific, *de dicto*, or non-presuppositional interpretation. Although such a strong correlation may be observed in certain parts of the grammar, it certainly does not hold across the board. We have seen that the association between case and a strong interpretation can be counteracted by the association of this case with an inherent argument feature (a split case alternation). This means that these approaches –although right in pointing out the existence of a correlation between case and certain interpretations– are too limited as they can only account for part of the data observed in (multidimensional) DOM systems.

Alternative accounts try to characterize the complete language-specific patterns of DOM by means of reference to a hierarchy (Bossong, 1985; Lazard, 1998; Aissen, 2003). The formally best worked-out proposal is that of Aissen (2003) who starts from the following simplex hierarchies:

(44) Human > animate > inanimate

(45) Pronoun > proper name > definite NP > indefinite specific NP > indefinite non-specific NP

In order to describe multidimensional DOM systems she crosses the two hierarchies which results in a large matrix ranking a set of composite properties (e.g. human pronoun, inanimate definite NP). Although certain property combinations are not ranked universally with respect to one another they have to assume a certain ranking in the language-specific characterization of DOM patterns. We believe that such a hierarchy-based approach is not viable for a number of reasons. First, the crossing of the two hierarchies in order to describe multidimensional DOM systems seems to assume that the two features play an equal role in such systems. However, as we have argued in the previous section, languages seem to give priority to one feature over the other and they may differ in their prioritizing. Thus, instead of fixing the ranking between universally unordered sets of composite properties, languages seem to apply one dimension after the other.

Secondly, this kind of approach seems to presuppose that all features involved in DOM are involved in the same way. However, as we have shown in the previous section some features may trigger the occurrence of overt case marking, whereas others are the result of the occurrence of overt case marking. The inability of hierarchy-based approaches to account for this variation becomes clearest in the case of fluid alternations where they often have to resign to analyses in terms of optionality. This is due to the fact that they cannot acknowledge that instead there is a change in the relation between the argument feature and the overt case marking.

On a more fundamental level, in our opinion hierarchies are not needed to state the language-specific patterns of DOM for the languages under discussion. These languages are fundamentally different from those that do make reference to such a hierarchy. For instance, in the Papuan language Awtuw, overt object marking is dependent on the ranking of the object with respect to the subject on an animacy hierarchy such that object marking only occurs when the object is at equal or higher rank than the subject (de Swart, 2007; Malchukov, 2008). This makes this type of language similar to the inverse type where verbal agreement is dependent on the ranking of the object with respect to the subject on a hierarchy. By contrast, in order to determine object marking in the languages under discussion in this article no reference has to be made to the subject. This seems to obviate the need for a hierarchy in the characterization of these languages.

In the previous section we have shown that the case-marking systems of the languages under discussion can be described in terms of

binary features and we take it that this approach can be extended to (multidimensional) DOM systems in general. From a descriptive point of view, the approach sketched in the previous section has (at least) two clear advantages. First, it allows for a representation which brings better to the fore the organization underlying multidimensional DOM systems. Secondly, it provides a more principled account of why meaning variation occurs in exactly those areas of the system where it occurs. Instead of relying on optionality mechanisms, meaning variation is only expected in those areas which are not taken up by split alternations. An approach based on features is much more flexible than one based on hierarchies and hence can deal both with hierarchical and anti-hierarchical or other 'unexpected' DOM patterns. In the absence of restrictiveness this kind of approach, however, has to rely on other principles in case certain patterns turn out to be more likely cross-linguistically than others. These issues are further discussed in Section 5. In the next section we present a formal model which accounts for the interaction between split and fluid marking.

4 A rule-based analysis of multidimensional DOM

The aim of this section is to provide an analysis of multidimensional DOM in a rule-based sign grammar formalism, in which the notion *sign of language L* is defined in terms of deriving a composite sign by applying a rule to some component signs. The analysis captures both the split as well as the fluid case alternations patterns observed in DOM. Importantly, in this analysis we characterize language-specific multidimensional DOM patterns without making reference to hierarchies.

Our rule-based analysis will be couched in a variant of sign grammar as defined in Kracht (2003, 181ff). The basic notions of this grammar formalism are that of *sign* as a form-category-meaning triple, and that of *rule*, which specifies (i) an operation on the formal entities of the component signs, (ii) an operation on the categories of the component signs, and (iii) an operation on the meaning of the component signs. Given a set BS of basic signs and a set of such rules G, a sign s_n is part of a language L iff (i) it is a basic sign, or (ii) it can be derived relative to BS and G. A sign s_n can be derived relative to BS and G iff (i) s_n is in BS, or (ii) there are signs $s_0, \ldots, s_{n-1}$ and an n-ary rule R such that $R(s_0, \ldots, s_{n-1}) = s_n$, and the signs $s_0, \ldots, s_{n-1}$ can be derived relative to BS and G (in finitely many steps). Categories are sets of pairs consisting of an attribute and a set of values. For

example, the attribute-value structure

$$\left[\begin{array}{lcl} \text{HUM} & : & \{+\} \\ \text{SPEC} & : & \{+,-\} \end{array}\right]$$

represents the set of signs referring either to specific or nonspecific humans. If the set of values for a feature F contains more than one value, then we say that F is underspecified, and if it contains all the values of a particular feature, then we say that the feature is fully underspecified. If the set of values for a feature F is a singleton set, we say that the value of F is specified. The set of values for a feature may not be empty. By convention, if the set of values contains all possible values for a feature, the feature may be omitted from the feature structure. Assuming that the feature SPEC has only the two values + and −, the following attribute value structure is equivalent to the previous one.

$$\left[\begin{array}{lcl} \text{HUM} & : & \{+\} \end{array}\right]$$

Given two categories C and C' we say that C is subsumed under C' iff for all features F in C it holds that the value set of F in C is a subset of the value set of F in C'. For example:

$$\left[\begin{array}{lcl} \text{CAT} & : & \{indef\} \\ \text{CASE} & : & \{pe\} \\ \text{SPEC} & : & \{+\} \end{array}\right]$$

is subsumed under the feature structure:

$$\left[\begin{array}{lcl} \text{CAT} & : & \{indef\} \\ \text{CASE} & : & \{pe\} \\ \text{SPEC} & : & \{+,-\} \end{array}\right]$$

If a rule requires a component sign of category C then it applies to all signs whose category is subsumed under C. Where no confusion arises we shall indicate a singleton set $\{a\}$ as a.

We will illustrate our approach through the case system of Romanian as outlined in Section 2.2 above. We first sketch our account in informal terms after which we will present a formalization. To analyse differential object marking in Romanian, we propose two rules for combining a direct object with a verb. The first rule combines a morphologically unmarked NP with the verb, provided the NP satisfies certain conditions C_1. The second rule combines a *pe*-marked NP with the verb, provided the NP satisfies a different set of conditions C_2. Split case marking is thus analysed by postulating that two rules impose *different* conditions on noun phrases. For example, the

condition that the NP be a pronoun is only part of the second rule, but not of the first rule, so that only the rule applying to *pe*-marked NPs can combine direct object pronouns. In other words, we have captured that *pe*-marking is obligatory for pronominal direct objects. On the other hand, indefinite inanimate NPs can only be combined by the first rule (the second rule requires indefinites to be human), so that indefinite inanimate direct objects are always morphologically unmarked. Thus by postulating two rules with different conditions on the type of NP it is possible to analyse the split marking of direct objects in Romanian.

The fact that indefinite human NPs may or may not be *pe*-marked is analysed by allowing both rules to apply to this type of NP. If both rules apply to indefinite animate NPs under exactly the same conditions (that is without stating any additional constraints on indefinite human NPs), then the result would be a 'free' alternation of *pe*-marked and morphologically unmarked indefinite human direct objects. In Romanian, however, the case alternation on indefinite human direct objects cannot be said to be free, since the presence of *pe*-marking implies that the NP is to be interpreted as specific. To analyse this, we assume that although both rules apply to indefinite human NPs, they do not do so under exactly the same conditions. The difference is that the second rule states that the indefinite human NP is specific, whereas the first rule is silent on this issue.

An important challenge for the analysis of the relation between overt case marking and the interpretation of indefinite human NPs is that this relation is asymmetric – the *pe*-marking of such an NP implies a specific interpretation, but the nominative marking does not imply a non-specific interpretation. Instead morphologically unmarked indefinite human direct objects can be either specific or non-specific. To analyse this fluid and asymmetric marking of indefinite human NPs, we assume that lexical items specify the values of some but not all features. For example they always specify the value for the feature HUM (standing for 'human'), but they never specify the value of the feature SPEC (standing for 'specificity'). In other words, HUM is an inherent lexical feature whereas SPEC is not. The fact that lexical items do not specify the value of the feature SPEC will be represented by means of SPEC : $\{+, -\}$, meaning that the value of the specificity feature is either specific $(+)$ or non-specific $(-)$. The rules combining direct objects can refer to both types of features. If a rule states that a non-inherent feature has a certain value, it essentially specifies information which has been left underspecified by the lexical item. So a feature value is the 'result' of applying a rule if (and only if) the rule specifies a value which has been left underspecified by the lexical

items. If on the other hand, the rule states that an inherent feature has a certain value, it cannot be said to specify this value, since by definition the value of an inherent feature is already specified by the lexical item. Instead it is the specification of this inherent feature on the lexical item that triggers the application of the rule. The asymmetric relation between case marking and specificity (*pe*-marking on indefinite humans implies specificity, but lack of case marking does not imply non-specificity) is then analysed as follows. The second rule applies to different types of NPs (as developed in detail below), one of them being indefinite, human and specific NPs. Note that the first two features are inherent, whereas the last one is non-inherent. This rule can therefore be said to specify the value 'specific' (+) for the feature SPEC, so that the result is a specific interpretation of the NP. The first rule also applies to different types of NPs, one of them being indefinite NPs. Since this first rule does not constrain the application to indefinite human NPs which are specific, the value of the feature SPEC remains underspecified (SPEC : $\{+, -\}$), and consequently the NP may be interpreted either as specific or as non-specific.

After having outlined the basic ideas for the analysis of split and fluid case marking we begin with the formulation of the two rules. What they have in common, is that they both assign the patient-like role of the verbal sign to the semantic value of the direct object sign. To account for this we postulate the same semantic operation for both rules, namely:

(46)
$$f^{SEM}(ARG, PRED(x, y)) = PRED(x, ARG)$$

For simplicity, we assume that the semantic value of the transitive verb sign is the Curried function $\lambda y.\lambda x.PRED(x, y)$ and that the semantic operation is functional application (FA):

(47)
$$FA(ARG, \lambda y.\lambda x.PRED(x, y)) = [\lambda y.\lambda x.PRED(x, y)](ARG)$$

We also assume that the two rules share the syntactic operation, so that the formal result of combining the direct object sign with the verbal sign is the concatenation of the direct object exponent to the right of the verb exponent.

(48)
$$f^{SYN}(x, y) = y \_ x$$

where x is the formal exponent of the direct object sign, and y is the formal exponent of the verb sign.

The categories are combined by means of an operation which results in the same category as the transitive verb sign (that of verbs), except that the value for the CAT attribute is *itv*.

(49) $f^{\text{CAT}}(\left[\ \text{CAT: } pro \sqcup name \sqcup def \sqcup indef\ \right], \left[\ \text{CAT: } tv\ \right]) = \left[\ \text{CAT: } itv\ \right]$

Putting these operations together we get the common core of both rules, namely:

(50) $R_{tr}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) =$
$= \langle f^{SYN}(e_1, e_2), f^{\text{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle =$
$= \langle e_2{}^\frown e_1, \left[\ \text{CAT: } itv\ \right], m_2(m_1)\rangle$

The range of component signs to which a particular rule applies is determined by stating the category of the component signs. Categories are analysed as sets of attribute value pairs. So the rules implementing for example the multidimensional DOM pattern in Romanian need to state (among other things) the following category information:

- the first rule combines a direct object with a transitive verb, provided that the indefinite NP is NOM-marked
- the second rule combines a direct object with a transitive verb, provided that the indefinite NP is *pe*-marked, specific and human

So the information that needs to be encoded in the direct object category of the first rule is that the sign is indefinite and marked as nominative. To implement these restrictions we need to distinguish two formal features, namely CAT and CASE, where CAT can have (among others) the values $pro, name, def, indef$ for nominal signs, and tv, itv for verbal signs, and CASE has (among others) the values *nom* and *pe*. In addition, we need to distinguish two semantic features, namely HUM with values $+, -$ and SPEC with values $+, -$.

The first rule can then be formulated as follows:

(51) $R_{tr_1}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) =$
$= \langle f^{SYN}(e_1, e_2), f^{\text{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle =$
$= \langle e_2{}^\frown e_1, \left[\ \text{CAT:}\quad itv\ \right], m_2(m_1)\rangle,$
where $c_1 = \left[\begin{array}{ll}\text{CAT:} & indef\\ \text{CASE:} & nom\end{array}\right]$ and $c_2 = \left[\ \text{CAT:}\quad tv\ \right]$

Note that since the feature SPEC is not mentioned in the attribute value structure of the nominal component sign, this is by our convention equivalent to having SPEC : $\{+, -\}$ in the feature structure,

which means that the rule applies to the set of all indefinite nominative NP signs, irrespective of their value for specificity. Therefore this rule applies to both specific and non-specific indefinite NPs, and thus captures the generalization that nominative indefinite NPs are not restricted to non-specific interpretations.

The second rule requires an indefinite NP to be both *pe*-marked and specific:

(52)
$$
\begin{aligned}
&R_{tr_2}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) = \\
&= \langle f^{SYN}(e_1, e_2), f^{\mathrm{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle = \\
&= \langle e_2{\frown}e_1, \left[\begin{array}{ll} \textsc{cat}: & \mathrm{itv} \end{array}\right], m_2(m_1)\rangle, \\
&\text{where } c_1 = \left[\begin{array}{ll} \textsc{cat}: & \mathit{indef} \\ \textsc{case}: & \mathit{pe} \\ \textsc{hum}: & + \\ \textsc{spec}: & + \end{array}\right] \text{ and } c_2 = \left[\begin{array}{ll} \textsc{cat}: & \mathit{tv} \end{array}\right]
\end{aligned}
$$

The application of the second rule has to be extended to pronouns, names and definite NPs denoting humans, in order to capture the fact that these NPs can also be object marked. The fact that they must be marked follows from the fact that they can be combined with the verb only by this rule. This extension can be achieved by adding a disjunct to the category of the direct object sign:

(53)
$$
\begin{aligned}
&R_{tr_2}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) = \\
&= \langle f^{SYN}(e_1, e_2), f^{\mathrm{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle = \\
&= \langle e_2{\frown}e_1, \left[\begin{array}{ll} \textsc{cat}: & \mathit{itv} \end{array}\right], m_2(m_1)\rangle, \\
&\text{where } c_1 = \left[\begin{array}{ll} \textsc{cat}: & \mathrm{indef} \\ \textsc{case}: & \mathrm{pe} \\ \textsc{hum}: & + \\ \textsc{spec}: & + \end{array}\right] \sqcup \left[\begin{array}{ll} \textsc{cat}: & \mathit{pro} \sqcup \mathit{name} \sqcup \mathit{def} \\ \textsc{case}: & \mathrm{pe} \\ \textsc{hum}: & + \end{array}\right] \\
&\text{and } c_2 = \left[\begin{array}{ll} \textsc{cat}: & \mathit{tv} \end{array}\right]
\end{aligned}
$$

However, since pronouns for inanimate entities are also marked, it is necessary to split the second disjunct into two disjuncts, one accounting for personal pronouns irrespective of the animacy of the entity they stand for, and the other accounting for names and definite NPs standing for humans:

(54)
$$
\begin{aligned}
&R_{tr_2}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) = \\
&= \langle f^{SYN}(e_1, e_2), f^{\mathrm{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle = \\
&= \langle e_2{\frown}e_1, \left[\begin{array}{ll} \textsc{cat}: & \mathit{itv} \end{array}\right], m_2(m_1)\rangle,
\end{aligned}
$$

$$\text{where } c_1 \begin{bmatrix} \text{CAT:} & indef \\ \text{CASE:} & pe \\ \text{HUM:} & + \\ \text{SPEC:} & + \end{bmatrix} \sqcup \begin{bmatrix} \text{CAT:} & pro \\ \text{CASE:} & pe \end{bmatrix} \sqcup \begin{bmatrix} \text{CAT:} & name \sqcup def \\ \text{CASE:} & pe \\ \text{HUM:} & + \end{bmatrix}$$

$$\text{and } c_2 = \begin{bmatrix} \text{CAT:} & tv \end{bmatrix}$$

Due to the disjunctive nature of this rule the ordering of its components cannot be used to derive the prioritizing visualized in the figures in Section 2 above. This prioritizing is, however, reflected in the amount of information specified in the attribute-value matrices of the disjuncts. In general, dominating alternations will contain less information than dominated ones.

Analogously, it is necessary to extend the application of the first rule also to cases in which names and definite NPs refer to animals or inanimate entities, resulting in:

(55)
$$R_{tr_1}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) =$$
$$= \langle f^{SYN}(e_1, e_2), f^{\text{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle =$$
$$= \langle e_2 \frown e_1, \begin{bmatrix} \text{CAT:} & itv \end{bmatrix}, m_2(m_1)\rangle,$$
$$\text{where } c_1 = \begin{bmatrix} \text{CAT:} & indef \\ \text{CASE:} & nom \end{bmatrix} \sqcup \begin{bmatrix} \text{CAT:} & name \sqcup def \\ \text{CASE:} & nom \\ \text{HUM:} & - \end{bmatrix}$$
$$\text{and}$$
$$c_2 = \begin{bmatrix} \text{CAT:} & tv \end{bmatrix}$$

These two rules account for the following case alternation in Romanian differential object marking (where + signifies the possibility of *pe*-marking, whereas − signifies the possibility of nominative marking, cf. also the figure in (31) above):

(56)

	pro	name	def	indef	
				+spec	-spec
human	+	+	+	±	–
non-human	+	–	–	–	–

Note, that in addition to allowing both nominative and *pe*-marking for human specific indefinite NPs (since both rules can apply for this type of direct object), the two rules capture (i) that a *pe*-marked indefinite can only be interpreted as specific (only the second rule can apply) and (ii) that a NOM-marked indefinite can be both specific and non-specific (due to the underspecification of the feature SPEC).

In sum, our framework accounts for the full range of patterns described in Section 2 above. However, as we have also discussed above, it the facts in Romanian and other multidimensional DOM systems

may turn out to be more complex. That is, other factors such as construction type, as discussed above, or verb class (see e.g. von Heusinger and Onea 2008) may be involved as well. Moreover, we have not yet accounted for the behavior of human definite NPs. We think that our formalism is flexible enough to allow for the characterisation of these facts too –once they have been fully cleared up. Each such factor will be incorporated as a restriction on the category information of a sign.

An important aspect of these rules is that they can be applied both in production and in understanding. A speaker who wants to express the thought that she wanted to read *War and Peace* may choose the phrase *un roman* ('a novel') in order to refer to this specific novel, for example because she is addressing a child who has not heard of *War and Peace*. Despite the specific novel that she has in mind, the category of the phrase *un roman* does not express this information.

$$\begin{bmatrix} \text{CAT:} & indef \\ \text{CASE:} & nom \\ \text{SPEC:} & \{+,-\} \\ \text{HUM:} & - \end{bmatrix}$$

The same holds for the choice of rule to express the grammatical function of a noun phrase. Despite the fact that I have a particular friend of mine in mind, it is not necessary to (i) use the phrase *pe un prieten* ('PE a friend') and (ii) apply the second (*pe*-marking) direct object rule, because not every aspect of my thought needs to be encoded. I can also (i) use the phrase *un prieten* and (ii) apply the first (nominative) direct object rule in order to express the idea that I wanted to visit a friend of mine, and thus leave the information that I mean a particular friend (as opposed to just any friend) unexpressed. Consequently, the noun phrase I use to refer to the particular friend of mine will be in the nominative case. Faced with a nominative NP my addressee cannot apply the second rule (which applies only to *pe*-marked NPs), and therefore has to use the first rule, which does not specify whether the NP is to be interpreted specifically or not.

To conclude, we repeat the three main properties of our rule-based analysis of DOM in Romanian. First, by postulating two direct object rules which apply to different (albeit overlapping) sets of NPs, we can explain the case marking split: if an NP can only be combined by the first direct object rule, then it is obligatorily nominative, whereas if it can only be combined by the second rule, it is obligatorily *pe*-marked. Secondly, the fluid case alternation on indefinite human NPs is captured by allowing both rules to apply to this type of NP. And thirdly, the difference between 'trigger' and 'result' properties is derived from the mechanism of (under)specification. The values of some features

(the so-called inherent features) are fully specified by the lexical items, whereas the values of other features (i.e. the non-inherent features like e.g. SPEC) are specified as a result of applying the second direct object rule. We have demonstrated our approach for Romanian, but it applies equally to the other languages discussed.Unfortunately, we cannot illustrate this due to space limitations.

5 Language description, abstraction and language comparison

An important point to note about the formalism within which this analysis has been expressed is that it does not prevent the formulation of DOM patterns which go against the cross-linguistic DOM generalization. For example, it is possible to formulate a pattern in which names, definite and indefinite NPs are ACC-marked whereas pronouns are morphologically unmarked. The following two rules characterize precisely this pattern:

(57) $R_{tr_3}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) =$
$= \langle f^{SYN}(e_1, e_2), f^{\text{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle =$
$= \langle e_2{\frown}e_1, \left[\begin{array}{ll} \text{CAT:} & itv \end{array}\right], m_2(m_1)\rangle,$
where $c_1 = \left[\begin{array}{ll} \text{CAT:} & pro \\ \text{CASE:} & nom \end{array}\right]$ and
$c_2 = \left[\begin{array}{ll} \text{CAT:} & tv \end{array}\right]$

(58) $R_{tr_4}(\langle e_1, c_1, m_1\rangle, \langle e_2, c_2, m_2\rangle) =$
$= \langle f^{SYN}(e_1, e_2), f^{\text{CAT}}(c_1, c_2), f^{SEM}(m_1, m_2)\rangle =$
$= \langle e_2{\frown}e_1, \left[\begin{array}{ll} \text{CAT:} & itv \end{array}\right], m_2(m_1)\rangle,$
where $c_1 = \left[\begin{array}{ll} \text{CAT:} & name \sqcup def \sqcup indef \\ \text{CASE:} & acc \end{array}\right]$ and
$c_2 = \left[\begin{array}{ll} \text{CAT:} & tv \end{array}\right]$

In the face of so much flexibility it is of course justified to ask whether the framework is not too liberal. Should the formalism not better be restricted so that such patterns cannot be characterized? The simple answer to such questions is that this flexibility is necessary, because this pattern of DOM is actually attested in Nganasan (Samoyedic) as discussed by Filimonova (2005). Likewise, our framework allows for the formulation of other 'unexpected' patterns such as that found in Hup (Makú; Epps 2005, 2008) in which plurality of nouns plays an important role. See also Bickel and Witzlack-Makarevich (2008) for

exceptions to hierarchical patterns.

But then which linguistic patterns of DOM should the grammar formalism exclude, if any at all? In view of such rare but attested patterns, it may be wiser not to exclude any logically possible DOM pattern by restricting the framework. This leads to a more general question: which linguistic patterns should the grammar formalism exclude? In line with Hawkins (2004) and others, we assume that unattested patterns need not be excluded by the grammatical framework as a matter of principle, if they may also be ruled out by factors which are external to the grammatical framework. Likewise, the fact that certain regularities seem to emerge when comparing DOM systems may ultimately be traced back to a historical origin as a pure disambiguation mechanism shared by such systems (see e.g. Haspelmath 1999; de Swart 2007).

In our view, cross-linguistic patterns of DOM are abstractions from language-specific patterns, which may or may not be universal. If, for example, the languages of a particular language family (or set of language families) show theoretically interesting similarities in their language-specific pattern of DOM, this justifies analyzing these similar but not identical patterns as instantiations of one abstract pattern (cf. Bickel and Witzlack-Makarevich 2008). Note that this does neither require nor imply that the abstract pattern has to be instantiated by all languages with DOM. This cross-linguistic abstract pattern need not be universal, it only needs to be of sufficient theoretical interest. Viewing things this way, it is not surprising that that different hierarchies have been proposed for different classes of languages to capture different cross-linguistic generalizations of a cluster of related patterns (for DOM compare for instance Bossong 1985; Lazard 1998 and Aissen 2003). So, comparative concepts in the sense of Haspelmath (2008a,b) do not need to be universal – they only need to be sufficiently abstract so that theoretically relevant generalizations can be formulated.

Finally, should hierarchies be used not just to formulate but also to explain cross-linguistic generalizations? If these hierarchies represent innate constraints on grammatical knowledge (as e.g. proposed by Kiparsky 2008), and there were no exceptions to the cross-linguistic pattern, then appeal to hierarchies may explain the cross-linguistic pattern. However, since cross-linguistic DOM patterns allow for exceptions, they cannot be explained solely by internal factors (innateness), but should be explained by an interaction between internal and external factors.

6 Conclusion

After discussing a number of multidimensional differential object marking patterns, we observed that the various argument properties are related to case marking in different ways. After discussing previous accounts of DOM, which typically characterize the language-specific patterns by reference to cross-linguistic animacy or referentiality hierarchies, we propose a rule-based analysis (couched in a sign grammar formalism) which characterizes DOM patterns without reference to such hierarchies. Importantly, the analysis also captures the different ways in which the argument properties relate to case, given that rules can impose e.g. a specific interpretation on an NP which would otherwise be left underspecified for the specificity value. Since only non-inherent properties can be left underspecified, we predict that only such properties can partake in fluid alternations. Finally, we show that the flexibility of the formalism is in fact necessary in order to characterize rare DOM patterns which provide exceptions to the cross-linguistic generalization about DOM.

Acknowledgements

We thank the two anonymous reviewers for their constructive comments which helped to improve this article. Furthermore we thank the participants of the Stuttgart Workshop on Case Variation for their comments. A special thanks to Klaus von Heusinger and Helen de Hoop with whom we discussed many of the issues presented in this article. This research was funded by the German Science Foundation (project C2 *Case and Referential Context* in the SFB 732 *Incremental Specification in Context*), which we gratefully acknowledge. Peter de Swart also received financial support from the Netherlands Organization for Scientific Research (NWO) (grants 360-70-220, *Animacy* and 275-89-003, *The Status of Hierarchies in Language Production and Comprehension*), which is hereby gratefully acknowledged.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory 21*(3), 435–483.

Bickel, B. and A. Witzlack-Makarevich (2008). Referential scales and case alignment: reviewing the typological evidence. In M. Richards and A. L. Malchukov (Eds.), *Scales*, Volume 86 of *Linguistische Arbeitsberichte*. University of Leipzig.

Bleam, T. (2005). The role of semantic type in differential object marking. *Belgian Journal of Linguistics 19*, 3–27.

Bossong, G. (1985). *Differentielle Objektmarkierung in den Neuiranischen Sprachen*. Tübingen: Narr.

Brugé, L. and G. Brugger (1996). On the accusative *a* in Spanish. *Probus 8*, 1–51.

Butt, M. (1993). Object specificity and agreement in Hindi/Urdu. In K. Beats, G. Cooke, D. Kathman, S. Kita, K. McCullough, and D. Testen (Eds.), *Papers from the 29th Regional Meeting of the Chicago Linguistic Society. Volume 1: The Main Session*, pp. 89–104.

Danon, G. (2001). Syntactic definiteness in the grammar of Modern Hebrew. *Linguistics 39*(6), 1071–1116.

Danon, G. (2006). Caseless nominals and the projection of DP. *Natural Language and Linguistic Theory 24*(4), 977–1008.

Delbecque, N. (2002). A construction grammar approach to transitivity in Spanish. In K. Davidse and B. Lamiroy (Eds.), *The Nominative/Accusative and Their Counterparts. Case and Grammatical Relations across Languages*, pp. 81–130. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Dobrovie-Sorin, C. (1994). *The Syntax of Romanian: Comparative Studies in Romance*. Berlin: Mouton De Gruyter.

Dobrovie-Sorin, C. (2007). Article-drop in Romanian and extended heads. In G. Alboiu, L. Avram, A. Avram, and D. Isac (Eds.), *Pitar Mos: A Building with a View. Festschrift Alexandra Cornilescu*, pp. 99–107. Bucharest: Bucharest University Press.

Enç, M. (1991). The semantics of specificity. *Linguistic Inquiry 22*(1), 1–25.

Epps, P. (2005). *A Grammar of Hup*. Ph. D. thesis, University of Virginia.

Epps, P. (2008). Hup's typological treasures: Description and explanation in the study of an Amazonian language. *Linguistic Typology 12*, 169–193.

Filimonova, E. (2005). The noun phrase hierarchy and relational marking: Problems and counterevidence. *Linguistic Typology 9*(1), 77–113.

García García, M. (2005). Differential object marking and informativeness. In K. von Heusinger, G. A. Kaiser, and E. Stark (Eds.), *Proceedings of the Workshop 'Specificity and the evolution/emergence of nominal determination systems in Romance'*, pp. 17–32. Konstanz: Universität Konstanz.

Guntsetseg, D. (2009). Differential Object Marking in Khalkha Mongolian. In R. Shibagaki and R. Vermeulen (Eds.), *Proceedings of the 5th Workshop on Formal Altaic Linguistics (WAFL5)*, Volume 58. MIT Working Papers in Linguistics.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft 18*(2), 180–205.

Haspelmath, M. (2008a). Comparative concepts and descriptive categories in cross-linguistic studies. Unpublished manuscript, Department of Linguistics, Max Planck Institute for Evolutionary Anthropology.

Haspelmath, M. (2008b). Descriptive scales versus comparative scales. In M. Richards and A. L. Malchukov (Eds.), *Scales*, Volume 86 of *Linguistische Arbeitsberichte*. University of Leipzig.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

von Heusinger, K. (2008). Verbal semantics and the diachronic development of differential object marking in spanish. *Probus 20*, 1–33.

von Heusinger, K. and G. Kaiser (2003). Animacy, specificity, and definiteness in Spanish. In K. von Heusinger and G. Kaiser (Eds.), *Proceedings of the Workshop "Semantic and Syntactic Aspects of Specificity in Romance Languages"*, pp. 41–65. Universität Konstanz: Arbeitspapier 113.

von Heusinger, K. and J. Kornfilt (2005). The case of the direct object in Turkish: Semantics, syntax and morphology. *Turkic Languages 9*, 3–44.

von Heusinger, K. and E. Onea (2008). Triggering and blocking effects in the diachronic development of dom in romanian. *Probus 20*(1), 67–110.

de Hoop, H. (1992). *Case Configuration and Noun Phrase Interpretation*. Ph. D. thesis, Groningen University. Groningen Dissertations in Linguistics 4.

de Hoop, H. and A. Malchukov (2007). On fluid differential case marking: A bidirectional OT approach. *Lingua 117*(9), 1636–1656.

de Hoop, H. and B. Narasimhan (2005). Differential case-marking in Hindi. In M. Amberber and H. de Hoop (Eds.), *Competition and Variation in Natural Languages: The Case for Case*, pp. 321–345. Amsterdam: Elsevier.

Hopper, P. J. and S. A. Thompson (1980). Transitivity in grammar and discourse. *Language 56*(2), 251–299.

Jäger, G. (2003). Learning constraint sub-hierarchies: The bidirectional gradual learning algorithm. In R. Blutner and H. Zeevat (Eds.), *Pragmatics in OT*, pp. 251–287. Houndmills: Palgrave MacMillan.

Kachru, Y. (2006). *Hindi*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kiparsky, P. (2008). Universals constrain change; change results in typological generalizations. In J. Good (Ed.), *Language Universals and Language Change*, pp. 23–53. Oxford: Oxford University Press.

Kracht, M. (2003). *Mathematics of language*. Studies in Generative Grammar No. 63. Berlin: Mouton de Gruyter.

Lazard, G. (1998). *Actancy*. Berlin: Mouton de Gruyter.

Leonetti, M. (2004). Specificity and object marking: The case of Spanish *a*. *Catalan Journal of Linguistics 3*, 75–114.

Lidz, J. (1999). The morphosemantics of object case in Kannada. In *Proceedings of West Coast Conference on Formal Linguistics 18*. Cambridge, MA: Cascadilla Press.

Lidz, J. (2006). The grammar of accusative case in Kannada. *Language 82*(1), 10–32.

Malchukov, A. (2008). Animacy and asymmetries in differential case marking. *Lingua 118*(2), 203–221.

McGregor, R. (1995). *Outline of Hindi Grammar. Third edition*. Oxford: Oxford University Press.

Mohanan, T. (1990). *Arguments in Hindi*. Ph. D. thesis, Stanford University.

Primus, B. (2009). Animacy, generalized semantic roles, and differential object marking. Unpublished manuscript, University of Cologne.

Ramchand, G. (1997). *Aspect and Predication*. Oxford: Oxford University Press.

Singh, M. (1994). Thematic role, word order, and definiteness. In M. Butt, T. H. King, and G. Ramchand (Eds.), *Theoretical Perspectives on Word Order in South Asian Languages*, pp. 217–236. Stanford: CSLI Publications.

Stark, E. (2008). What is marked by differential object marking in romance. Paper presented at the workshop on Case Variation, University of Stuttgart, 19 June 2008.

de Swart, P. (2007). *Cross-linguistic Variation in Object Marking*. Ph. D. thesis, Radboud University Nijmegen. LOT Publications.

de Swart, P. and H. de Hoop (2007). Semantic aspects of differential object marking. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung 11*, pp. 568–581. Barcelona: Universitat Pompeu Fabra.

Torrego, E. (1998). *The Dependencies of Objects*. Cambridge, MA: The MIT Press.

Titel	Case and Referential Properties
Verfasser	Klein, Udo & de Swart, Peter
Eingereicht bei	von Heusinger, Klaus & de Hoop, Helen (Eds.) Lingua: Special Issue „Case Variation“
Eingereicht am	12.02.2009
Reviews erhalten am	09.06.2009
Überarbeitung wiedereingereicht am	20.07.2009
gegebenenfalls Zusage vom Serieneditor am	
Zusage zum Druck am	

To: "Peter de Swart" <P.J.F.de.Swart@rug.nl>
From: Klaus von Heusinger <Klaus.vonHeusinger@ling.uni-stuttgart.de>
Subject: Re: Revision Volume on Case Variation
Cc: Helen, udo
Bcc:
X-Attachments:

Dear Peter and Udo,

thanks for the revised version of your paper and the list of corrections. I think you did a very good job and that the paper is now much better.

Helen pointed out to me that the final decision of acceptance or not is made by the journal editors - but both the reviewers and the editors of the special issue recommend publication - so I think that this should work out.

We keep you informed about the further procedure (which might take some more time since we are still waiting for some papers)

Best,

Klaus

>Dear Helen and Klaus,
>
>Please find attached the revised version of our paper.
>We have submitted a .tex, .bib. and .pdf file (which accoriding to the Lingua documentation is allowed). If the elsevier typesetters plug in their own style the references should be formatted their way without any problems. The word count of our paper (copying the pdf into ms word) stops at 11,481. We have also included a .pdf file in which we specify the changes made.
>
>If you need anything else, please let uw know. Good luck with all the editorial matters.
>
>best,
>
>Udo and Peter
>
>
>Attachment converted: Macintosh HD:revision.zip (pZIP/pZIP) (0026B2B8)