

Differential object marking

# Table of Contents

<b>List of abbreviations</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 The functional-typological approach . . . . .	3
1.3 Markedness and prototypes . . . . .	7
1.4 Outline of the work . . . . .	10
<b>2 Theoretical foundations</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Transitivity . . . . .	11
2.3 The puzzle of direct object . . . . .	18
2.4 Approaches to direct objecthood . . . . .	21
2.5 Differential object marking: previous approaches . . . . .	25
2.5.1 The Distinguishing Approach . . . . .	25
2.5.2 The Indexing Approach . . . . .	35
2.5.3 The Information Structure Approach . . . . .	41
2.5.4 Secondary topic and DOM . . . . .	42
<b>3 A model for DOM and DOI</b>	<b>48</b>
3.1 Towards a model for DOM and DOI . . . . .	48
3.2 Markedness and DOM: a reappraisal . . . . .	55
3.3 DOM, DOI and diachrony . . . . .	60
<b>4 Parameters of analysis</b>	<b>61</b>
4.1 Introduction . . . . .	61
4.2 DOM and DOI: structural variation . . . . .	61
4.3 The parameters . . . . .	65
4.3.1 Animacy . . . . .	65
4.3.2 Definiteness, identifiability and specificity . . . . .	69
4.3.3 Activation . . . . .	71

4.3.4	Topicality . . . . .	72
4.3.5	Focality . . . . .	74
4.3.6	Dislocation and topicalisation . . . . .	76
4.4	The sample . . . . .	79
<b>5</b>	<b>A typology of DOM and DOI: animacy and topic-worthiness</b>	<b>81</b>
5.1	Introduction . . . . .	81
5.2	Afro-Asiatic . . . . .	81
5.3	Australian . . . . .	86
5.4	Austro-Asiatic and Austronesian . . . . .	90
5.5	Indo-European and Dravidian . . . . .	95
5.6	Creoles . . . . .	102
5.7	Papua New-Guinea . . . . .	105
5.8	Tibeto-Burman . . . . .	113
5.9	South America . . . . .	118
5.10	Central and North America . . . . .	123
5.11	Africa . . . . .	128
5.12	Conclusion . . . . .	130
<b>6</b>	<b>A typology of DOM and DOI: topicality and definiteness</b>	<b>134</b>
6.1	Introduction . . . . .	134
6.2	Afro-Asiatic . . . . .	135
6.3	Nilo-Saharan . . . . .	149
6.4	Niger-Congo . . . . .	156
6.5	Indo-European and Dravidian . . . . .	161
6.6	Altaic . . . . .	168
6.7	Uralic . . . . .	173
6.8	Yukaghir . . . . .	178
6.9	South America . . . . .	180
6.10	North and Central America . . . . .	189
6.11	Austro-Asiatic, Austronesian and Papuan . . . . .	199
6.12	Tibeto-Burman . . . . .	210
6.13	Conclusion . . . . .	217
<b>7</b>	<b>DOM in serial verb constructions</b>	<b>219</b>
7.1	Introduction . . . . .	219
7.2	Sinitic . . . . .	220
7.3	Benue-Congo . . . . .	229
7.4	Conclusion . . . . .	232

<b>8</b>	<b>A case study: DOM and topicality in Romance</b>	<b>235</b>
8.1	Introduction . . . . .	235
8.2	DOM in Romance: some general remarks . . . . .	235
8.3	Italian and Italian dialects . . . . .	236
8.3.1	Northern and Standard Italian . . . . .	237
8.3.2	Gallo-Italian dialects . . . . .	245
8.4	DOM in Catalan, Portuguese, and French varieties . . . . .	250
8.4.1	Catalan and Portuguese . . . . .	250
8.4.2	French varieties . . . . .	256
8.5	DOM and topicality in Spanish and Sicilian . . . . .	259
8.5.1	Sicilian . . . . .	259
8.5.2	Spanish . . . . .	262
8.6	Conclusion . . . . .	265
<b>9</b>	<b>Conclusion</b>	<b>268</b>
	<b>Appendices</b>	<b>272</b>
<b>A</b>	<b>Languages in the sample</b>	<b>273</b>
<b>B</b>	<b>Main sources of information on the languages in the sample</b>	<b>278</b>
<b>C</b>	<b>List of the constructions analysed</b>	<b>283</b>
<b>D</b>	<b>Questionnaires</b>	<b>289</b>
D.1	Questionnaire 1 . . . . .	289
D.2	Questionnaire 2 . . . . .	290
	<b>Bibliography</b>	<b>324</b>

# List of Figures

3.1	Mapping between semantic and morphosyntactic properties of direct objects in DOM . . . . .	50
3.2	Mapping between semantic and morphosyntactic properties of direct objects in DOI . . . . .	52
3.3	Mapping between DOM and DOI, semantics and information structure . . . . .	54

# List of Tables

2.2	Transitivity parameters (Hopper and Thompson 1980: 252) . . . . .	13
2.4	The Individuation Hierarchy (Timberlake 1977: 162) . . . . .	13
2.6	Transitivity scale (Malchukov 2006: 333) . . . . .	16
3.2	Features of direct objects . . . . .	49
5.1	Distribution of DOM and DOI in Palauan . . . . .	93
5.2	Abui object prefixes . . . . .	109
6.1	Distribution of DOM in Sahidic Coptic . . . . .	138
6.2	Distribution of DOM in Komi dialects . . . . .	178
6.3	Object encoding in Kolyma Yukaghir (adapted from Maslova (2003b: 89) . . . . .	178
6.4	Distribution of case marking based on information structure in Tucanoan languages . . . . .	184
6.5	Distribution of case marking in Tariana . . . . .	185
6.6	Distribution of object markers in Galo . . . . .	211
A.1	Genetic affiliation of the languages in the sample . . . . .	277
B.1	Main bibliographical sources for the languages in the sample . . . . .	282
C.1	Summary of the constructions found in the languages of the sample	287

I

# List of abbreviations

The glossing conforms generally to the Leipzig Glossing Rules (see <http://www.eva.mpg.de/lingua/resources/glossing-rules.php> for details). Non-standard abbreviations are listed below. Language examples have been presented in standard orthography or in the orthography used by the source.

ADVR	adversative	AFFMT	affirmative
ASP	aspect	ASS	asseverative clitic
ATT	attributive	CNTR	contrastive
CONCESS	concessive	CP	completive
DEP	dependent marker	DIM	diminutive
DLMT	delimitative	DOM	differential object marking
DS	different subject	DYNAM	dynamic
FIN	finite marker	FV	final vowel
HIST	historical past	HUM	human
ILL	illative	IMPRS	impersonal
INCH	inchoative	INDEP	independent
INT	interrogative	ITG	intangible demonstrative
LK	linker	LOCUT	locutor marker
NF	non-feminine	NH	non-honorific
NHYP	non-hypothetical	NMLZ	nominaliser
NONLOCUT	non-locutor person marker	OBJ	object indexation
PART	partitive	PC	pronominal clitic
PDET	proximate determiner	POL	polite
POS	epistemic possibility	PREF	prefix
PRT	particle	REAL	realis
RED	reduplication	REP	reportative
SAP	speech act participant	SEQ	sequential
SGV	singulative	SS	same subject
STI	stance marker	SUB	subordinator
SUBJ	subject	SUP	superlative
TAM	TAM marker	THEME	thematic
UNM	unmarked		





# Chapter 1

## Introduction

### 1.1 Overview

Differential Object Marking (henceforth, DOM) and Differential Object Indexation (DOI) have been the object of a growing interest to linguist over the past decades, particularly in functional typology (cf. Bossong 1985, 1991, 1998, Comrie 1979, 1989, Croft 1988, among others)<sup>1</sup>. Basically, DOM and DOI are a variation in the encoding of direct objects, whereby only a subset of direct objects receives overt coding (DOM), or is indexed on the verb (DOI), depending upon semantic or pragmatic features of the direct object, such as animacy, definiteness, and specificity. Examples (1.1) from Sicilian (Indo-European, Romance) illustrate DOM, whereas examples (1.2) from Alaaba (Afro-Asiatic, Cushitic) illustrate the presence of indexation with animate direct objects:

(1.1) Sicilian (Indo-European, Romance)

(a) *Ciercu*            *a*    *u*    *cammarieri*  
search.1SG.PRS    DOM    DET    waiter  
"I'm looking for the waiter"

(b) *Ciercu*            *u*    *cammarieri*  
search.1SG.PRS    DET    waiter  
"I'm looking for the waiter" (personal knowledge)

(1.2) Alaaba (Afro-Asiatic, Cushitic)

---

<sup>1</sup>Following Croft 2003: 34), in this work I will use the term *indexation* instead of agreement, because indexation does not imply the presence of an overt NP which controls the distribution of the markers on the verb.

- (a) *ʔís(i)*            *ʔissát(a)*            *hikka*            *b-á*  
 3SG:M:NOM PRON3PL:ABS DEM2:ABS place-TN:M:ABS  
*soh-yó-ss(a)*  
 send-3SG:M:PFV-3PL.PC  
 "He sent them there (Schneider-Blum 2007: 74)
- (b) *mancú*                    *c'ub-éen(i)*  
 person.SGV:M:ABS knife-TN:F:LOC/INSTR  
*sh-éema-s(i)*  
 kill-3SG:POL:PFV-PC.3SG:M  
 "Somebody killed a man with a knife" (Schneider-Blum 2007: 142)

From a conceptual or "real-world" perspective, the event represented by the two clauses in (1.1) is essentially identical, namely that there is someone who is looking for a waiter. However, this very same event is differently conceptualised from a linguistic point of view: while the direct object in (1.1a) is overtly coded by the preposition *a*, the direct object in (1.1b) is left uncoded. This is due to the fact that the referent of the direct object in (1.1a) is identifiable and active in the current discourse, while the referent of the direct object in (1.1b) is not.

Traditionally, the contrast between overtly coded and uncoded forms are explained by differences in semantic properties of the direct object referent. Nonetheless, there seems to be no visible difference in the examples given in (1.1) with respect to the semantic properties of the NP. Differences in case-marking of direct objects have raised a number of questions among linguists since the 19th century (cf. for instance Diez 1863, Meyer-Lübke 1900). First, there has been an intense discussion as to the transitivity of the sentence containing the overtly coded direct object as opposed to the uncoded one. Second, the role of semantic and pragmatic properties like definiteness and animacy has been the topic of prolonged debate amongst linguists of different theoretical persuasions. While some scholars have argued for a key role of animacy and definiteness in shaping these constructions, others have considered them as epiphenomenal. Another point of discussion, especially in the past few years, has been the influence of information structure properties and discourse factors in DOM-systems (Nikolaeva 2001, Dalrymple and Nikolaeva 2011, Klumpp 2009, Shain and Tonhauser 2010). Fourth, the diachrony of DOM is an issue that is gaining greater prominence amongst researchers, even though there are only a few studies devoted to the diachronic development and stabilisation of DOM (see Pensado 1995, von Heusinger and Kaiser 2011). Finally, although DOM is well attested cross-linguistically, and has been thoroughly investigated in individual languages or language families, a comprehensive cross-linguistic study is still missing.

This work is a typological study of DOM and DOI based on a sample of 175 languages. The results of the typological study are complemented by a case-

study on the development of DOM in some Romance languages. Here I will deal with the cross-linguistic distribution of DOM and DOI, with particular regard to the role of referential parameters and information structure in shaping this distribution. In particular, attention is focused on (i) the relationship between the morphosyntactic and semantic properties of DOM and DOI systems, their diachronic sources and their discourse functions, and (ii) the similarities and differences in the properties and use of DOM on the one hand and DOI on the other. I show that, cross-linguistically, DOM and DOI are recurrently associated with two major construction types, topicalisations and dislocations. Diachronic data from genetically and geographically diverse languages show that DOM and DOI systems that appear to be based on animacy and definiteness synchronically are a result of the grammaticalisation of earlier topicality-based systems. In addition, in many languages, object markers are part of polysemy patterns involving (i) frame-setting expressions, such as spatio-temporal expressions (as in many Indo-European, Afro-Asiatic, Tucanoan, and Altaic languages), and (ii) topic and conditional markers (as in some Tibeto-Burman, Nilo-Saharan, and Austronesian languages). In some cases, even if the connection between object markers and topic markers is no longer transparent synchronically, it can still be reconstructed based on diachronic evidence. As will become clear in the next chapters, this means that, contrary to what has been often assumed in the literature, DOM systems do not primarily arise from the need to distinguish between the two arguments of a transitive clause, or to indicate a high degree of transitivity of the clause. Rather, these systems typically originate from the grammaticalisation of topic constructions, and they involve animate and definite direct objects because these particular types of direct objects, as opposed to others, are more likely to function as topics. Another major result of my study is that, although DOM and DOI systems seem to be based on the same parameters, they actually perform different discourse functions. DOM is often used to signal topic discontinuities, such as topic shifts and topic promotions, while DOI is used to maintain topic-continuity, in that it is consistently associated with highly accessible and continuous referents.

## 1.2 The functional-typological approach

The approach to DOM and DOI outlined in section 1.1 assumes that the notions of DOM and DOI should be defined so as to ensure cross-linguistic validity and applicability. This is one of the basic tenets of functional typology, which is the approach on which this work is based. The basic sense of typology involves the classification of languages based on structural properties, such as the order of main constituents in sentences. However, the distinguishing features of typology compared to other approaches to language are the emphasis placed on cross-linguistic

comparison and functional motivations of the structures found across languages. Typologists seek to discover patterns of grammatical variation found in the world's languages by means of cross-linguistic comparison. This method allows linguists to discover what is attested and what is not in languages and formulate principles about that variation which can be reasonably supposed to be universally valid. These principles, called universals, work like a sort of empirically-based constraints that predict the distribution of a given feature across languages. There are two main types of universals: absolute and implicational universals. Absolute, or unrestricted, universals are valid unexceptionally for all the languages with respect to a given feature, such as the presence or absence of vowels. However, only few absolute universals have been discovered so far. Implicational universals are much more common cross-linguistically. They describe a "restriction on logically possible language types that limits linguistic variation but does not eliminate it" (Croft 2003: 53) and are based on the relationship between at least two logically independent parameters. The label "implicational" derives from the fact that these universals are expressed as logical implications, such as "if a language has the feature  $x$ , then it has the feature  $y$ ". One of the simplest and best-known implicational universals predicts that if a language has nasal vowels, then it has oral vowels: therefore, there could be languages with oral vowels only or with both oral and nasal vowels. According to this implicational universal, no language is however found that does have nasal vowels but no oral vowels. Implicational universals provide a useful means to describe and constraint variation. Even though languages may show a huge amount of variation, such variation is predicted to conform to the restrictions established by the universal principles.

As is clear from the brief discussion above, typological investigations presuppose cross-linguistic comparison. Therefore, the issue of making a sample as representative as possible has been deeply discussed amongst typologists (see Dryer 1989, Rijkhoff et al. 1993, among others). There are different sampling procedures. A "variety sample" is a sample "intended to maximise the likelihood of capturing all the linguistic diversity for the phenomenon under study" (Croft 2003: 21). A "probability sample" consists of a randomly chosen subset of the world's languages designed to allow one to make statistical generalisations. However, the sample used in this study is not a statistically balanced one, due to the availability of good data and reliable grammatical descriptions. Rather, a convenience sample of 175 languages from all over the globe has been built, covering more than 25 language families, plus some creoles and isolates.

In turn, cross-linguistic comparison presupposes cross-linguistic comparability of phenomena across languages. The issue of cross-linguistic comparability is crucial to the functional-typological approach to language. Since languages vary to a great extent, it may turn out very difficult to identify a particular phenomenon on the ground of formal similarity. For instance, if the notion of subject

is defined only on the basis of purely structural criteria, such as the presence of case marking or indexation, we could be unable to identify subjects in languages that do not fulfil any of these criteria (Croft 2003: 14-15). Therefore, typologists make use of external parameters, semantically and/or pragmatically based, to define a phenomenon, be it syntactic, morphological or phonological. The assumption underlying this approach is that basic meanings (such as time or space) and communicative needs and skills (such as the need to identify referents), as well as perception and learning processes are shared by all human beings and therefore universal. A functional definition of linguistic phenomena allows linguists to deal with the enormous variation found in the world's languages without running the risk of comparing very different phenomena. This study takes a functional-typological approach, insofar as it relies upon functional factors to account for the cross-linguistic distribution of DOM and its diachronic development. Yet, as observed by Croft (2003: 15 ff.), the use of structural criteria could prove very useful in cross-linguistic analysis. Indeed, only few typological studies hinge upon purely functional criteria, being more common the use of mixed functional-formal criteria, which Croft (2003) defines as "derived structural definitions". The definition of the object of our analysis, namely what DOM and DOI are and how they can be identified, will be based mainly on functional criteria, even though some morpho-syntactic criteria will be used to carry out our cross-linguistic investigation of DOM and DOI.

As I have already mentioned, restrictions on possible language types are explained on the basis of functional motivations, since language structures are motivated both by their external functions (such as the use of language in discourse) and by cognitive and physiological constraints (such as memory loads). Two typical instances of functional principles widely used in typological studies are iconicity and economy. Iconicity can be roughly defined as the tendency of language structures to reflect the perception of the world by the speaker. To put it another way, linguistic structures are shaped by the perspective speakers have of the world (Croft 2003:102). Iconicity has two different facets, namely isomorphism and diagrammatic iconicity. Diagrammatic iconicity is based on the relationship between parts of the linguistic structures that "diagrammatically" mirrors a relationship between parts of the concepts that those expressions encode (Haiman 1983: 11). Isomorphism, or "one-form-to-one meaning" is a one-to-one correspondence between between parts of linguistic structure and parts of the conceptual structure (Croft 2003: 102 ff.). While iconicity "aims at making linguistic structures as transparent as possible" (Cristofaro 2003: 9), economy is based on the principle of "minimal effort", as it aims at making linguistic structures as simple as possible, by minimising linguistic expressions. Two different kinds of economy have been recognised by many scholars, paradigmatic and syntagmatic economy (Haiman 1983). Paradigmatic economy refers to the tendency for linguistic systems to

keep the number of signs as low as possible. Syntagmatic economy, on the other hand, refers to the tendency to simplify linguistic expressions either by reducing phonetic substance (as predicted by Zipf's law) or by avoiding to code explicitly unnecessary or context-inferable information. The relation between iconicity and economy has been the main topic of concern of many studies (Haiman 1983, Dubois 1987; Bybee 2001, among others), where it has been often argued that these two motivations can be viewed as competing forces which cannot be simultaneously fulfilled, i.e. one motivation will usually prevail over another. For instance, alignment systems, such as ergative vs. nominative systems, have been argued to be determined by two different competing motivations. For instance, Dubois (1987) argues that arguments in ergative languages (i.e. intransitive subject/direct object vs. agents) are grouped according to their "newness" in discourse. Arguments that are usually exploited to introduce new participants are put together. On the contrary, in nominative systems, arguments are grouped based on the semantic criterion of agentivity. In this case the more agentive participants, i.e. intransitive subjects and transitive agents, are aligned together. Similarly, case marking has been repeatedly argued to be determined by the two competing motivations of iconicity and economy. In economically motivated instances of case, the use of case marking will be as minimal as possible because of the "minimal effort" principle: case marking thus will be restricted only when there is a need to clearly distinguish participant. On the contrary, when iconicity is at work, case marking would be used more extensively, since iconicity requires linguistic expressions to be clearly and unambiguously coded. Therefore, while economy tries to keep linguistic expressions as simple as possible, avoiding the use of unnecessary overt marking, iconicity requires linguistic expressions to overtly code as much information as possible. As we will see in the next chapters, economy and iconicity, viewed as competing forces, have played a major role in the analysis of DOM systems, especially in recent work cast within the Optimality Theory framework (Aissen 2003).

The idea that the organisation of linguistic systems is determined by communicative needs makes the functional-typological approach crucial for the study of language change as well. The constraints on synchronic variation represented by typological hierarchies and implicational universals can be easily taken as representing grammaticalisation chains (cf. Cristofaro 2009, 2010; Croft 2003, ch. 2). In this view, which Croft has called "the dynamicisation of typology", synchronic and diachronic variation are not to be considered as two different and separate kinds of variation: synchronic variation is instead to be interpreted as a signal of a diachronic change in action. This idea will be especially useful to account for variation found in DOM and DOI systems cross-linguistically.

### 1.3 Markedness and prototypes

Two central explanatory concepts in functional typology, especially for the study of DOM and DOI, are the notion of markedness and the so-called prototype theory. We will discuss them briefly. As is well known, the concept of markedness was first developed to account for phonological phenomena within the *Cercle Linguistique de Prague*. Subsequently Greenberg (1966) adapted it to other levels of linguistic analysis and first utilised it in typology. Croft defines typological markedness as "a network of relationships among cross-linguistic asymmetrical patterns in grammar" (Croft 2003: 87). Basically, a markedness relation involves an asymmetry with respect to "grammatical properties of linguistic structures of otherwise equal linguistic elements" (Croft 2003: 87). Typological markedness, as opposed to the classic notion of markedness, concerns how a given conceptual category (i.e. a function) is encoded across the world's languages, and not how a language particular property of a particular grammatical category is encoded within a single language (Croft 2003: 88). That is, given a conceptual category  $x$ , typological markedness provides a means to compare different formal structures and discover interesting generalisations about how a given category is encoded cross-linguistically. Typological markedness patterns can be analysed as implicational universals, in that they provide constraints on possible language types. For instance, the structural criterion predicts that the marked value of an opposition (say, the accusative in nominative-accusative systems) exhibits as many morphemes as does the unmarked value (i.e. the nominative). This criterion excludes -at least in principle- a situation in which the nominative is, for instance, overtly coded, and the accusative is left uncoded. Synchronic markedness patterns can be exploited to predict diachronic changes as well. For instance, the generalisation about the markedness of nominative vis-à-vis accusative described above predicts that a zero accusative marker will not emerge unless a zero nominative marked emerges.

Over the years, a number of criteria to identify markedness patterns have been proposed in the literature: Croft (2003) grouped the various criteria first identified by Greenberg (1966) into four types: structural, inflectional, distributional (the latter two subsumed under the label of behavioural potential), and frequency criteria:

- Structural criterion: the marked member of an opposition exhibits as many morphemes as does the unmarked member;
- Inflectional criterion: the marked member of an opposition exhibits as many or less inflectional distinctions than the unmarked member;

- Distributional criterion: the marked member of an opposition occurs in as many or less grammatical environments than the unmarked member;
- Frequency criterion: the marked member of an opposition is as much or less frequent than the unmarked member.

Typological markedness patterns might be also viewed as instantiations of prototypical categories. Prototype theory has been developed first in psychology and cognitive science by Eleanor Rosch and her associates in the 70ies (cf. Rosch 1978). Subsequently, prototype theory has been incorporated in cognitive linguistics and functional-typological linguistics, where it has become a crucial device for analysing linguistic data. The central idea underlying prototype theory is that categories (here, linguistic categories) are not discrete entities, characterised by a set of properties shared by all members of a given category. In this model, which is deeply rooted in Aristotle's works on *Categories*, categories have to be clearly defined, and an entity can belong only to a single category. Prototype theory challenged this view, postulating that categories are indeed fuzzy and membership to a given category is a matter of degree (Rosch 1978, Lakoff 1987). Each category is defined by a set of core properties: some members share all or most of the core features of the category prototype, and therefore are more central. Members which lack core features but still share some of them with the prototype are said to be peripheral members of that category prototype. An often cited example of a prototype category is the "bird" category. A prototypical bird can be defined as an entity that shows some core features, such as feathers, beak, ability to fly, wings. Therefore, a sparrow or a robin can be considered as more prototypical members of the "bird" category than, so to speak, a penguin. As is clear, categories are thus graded, and become definable in terms of similarity to a central exemplar. The more core features an entity shows, the closer it gets to the prototype (Ghia 2010).

The application of prototype theory to grammatical concepts and categories is closely related to the notion of markedness. As noted by Croft (2003: 163 ff.), the evidence used to identify markedness patterns can be used to establish whether a member of a particular grammatical category is core or peripheral. First, the less marked member of a category (i.e. the member that show zero or minimal coding) is often deemed as the more central member of that category (structural criterion). Second, a central member of a category is expected to show more inflectional distinctions than a peripheral one (inflectional criterion). Third, a central member of a category is expected to occur in more contexts than a peripheral one (distributional criterion). Finally, a central member of a category is expected to be textually more frequent than a peripheral one (frequency criterion). In addition to these criteria, the marked member of an opposition will also show greater cognitive or semantic complexity than its unmarked counterpart.



Croft (2003: 164) further notes that a different combination of features can result in different prototypes: that is, there could be more than one unmarked combination of features. For instance, what is the marked combination of features in a prototype category may form the unmarked combination of another prototype. This phenomenon, called *markedness reversal*, has been invoked to account for various grammatical facts. For instance, while voicing seems to be the marked feature for obstruents, it is the unmarked feature for vowels and nasals (Croft 2003: 164, quoting Greenberg 1966).

As is widely accepted in the linguistic literature, the explanatory strength of the notion of markedness relies upon the interrelation between its different parameters. Thus, a more complex category will be less frequent than the less complex one, as well as less complex semantically and cognitively. Nonetheless, the notion of markedness has been extensively challenged in various studies. As a matter of fact, as noted by Haspelmath (2006) and Bybee (2010), the notion of markedness has been expanded and used in so many contexts that it has lost its explanatory power. For this reason, Haspelmath (2006) proposed to excise the notion of markedness as an explanatory concept in linguistics. In his view, markedness should be replaced by clearer and more substantive notions, such as "phonetic difficulty" or "frequency in texts" (Haspelmath 2006: 3; 30). Likewise, Blevins (2004) and Hume (2004) argue that markedness relations are epiphenomena of other factors, such as frequency or predictability and hence need not be present in linguistic theory.

While I find Blevins' and Hume's point of view rather drastic, I agree with Haspelmath and Bybee that markedness has been overused in linguistics as an overarching explanatory concept, thus losing its descriptive and explanatory value. For this reason, I have decided not to employ the terms "marked-unmarked" in this work. Rather, as I will discuss in Chapter 3, I will make direct reference to the concepts of frequency and expectedness to refer to the putative semantic and pragmatic markedness of topical, animate, and definite direct objects, while I will use the terms "uncoded-overtly coded" to refer to the presence of morphological marking.

In Chapter 2 we will see that works on DOM have largely made use of the concepts I have briefly discussed in this section, such as markedness, iconicity, economy, and prototypicality. Indeed, one of the main aims of this study is to evaluate the actual role played by these concepts. I will show that, while structural markedness effects are real and deserve explanations, there is no strong evidence for the cognitive or semantic marked status of topical direct objects in determining the appearance and development of DOM and DOI systems. Rather, these phenomena are better explained by making direct reference to notions like frequency and expectedness, as well as the functions fulfilled by DOM and DOI systems in discourse.

## 1.4 Outline of the work

This work is organised as follows. Chapter 2 sets out the theoretical foundations of this work: previous approaches to DOM and DOI are discussed along with some key notions pertaining to the study of DOM and DOI, such the notions of direct object and transitive clause. In chapter 3, a functional model of DOM and DOI systems is proposed, based upon the data presented in this study as well as upon the analysis of the problems found in the previous literature. Chapter 4 describes the sample used in this study and the morpho-syntactic, semantic, and information-structural parameters that will be taken into account in examining the cross-linguistic distribution of DOM and DOI systems.

In Chapters 5-7, the data concerning the cross-linguistic distribution of DOM and DOI are presented and thoroughly discussed. Chapter 5 deals primarily with DOM and DOI systems in which animacy synchronically takes priority over the other parameters. I show that DOM and DOI systems based on animacy are the result of the diachronic extension and conventionalisation of the constructions from topical direct objects to direct objects that show features usually associated with topics, like animacy. Chapter 6 analyses DOM and DOI systems primarily based on topicality. In this chapter, I adduce diachronic and synchronic evidence for the key role of topicality in determining the development and distribution of DOM and DOI systems cross-linguistically. Chapter 7 examines the distribution of DOM systems deriving from serial verb constructions. I show that this kind of DOM systems is governed by topicality as well. Chapter 8 is devoted to the analysis of the rise and distribution of DOM in Romance languages, with particular attention to less-studied Romance languages and dialects. Here we will again see close ties between the emergence of DOM systems and information structure. Finally, Chapter 9 summarises the major findings of this research and briefly discusses some further theoretical implications of these findings.

Three appendices present the data supporting the analysis. Appendices A-C provide information about the sources of information for the languages of the sample, their genetic affiliation, and geographic location. Appendix D presents the questionnaires used to collect the data analysed in Chapter 8 on Romance languages.

# Chapter 2

## Theoretical foundations

### 2.1 Introduction

In this chapter, I will first present some key concepts pertaining to the study of DOM and DOI. In particular, the notion of transitivity, as defined in functional-typological studies, will be discussed, since differences in object marking have been said to affect both formal and semantic transitivity. Subsequently, the notion of direct object will be illustrated. It will be argued that it is not possible to single out a cross-linguistically valid definition of direct object. Instead, a 'comparative concept' of a direct object (Haspelmath 2010) for cross-linguistic comparison will be proposed before discussing the notion of DOM. Finally, I will discuss and critically examine some previous approaches to DOM, which will be respectively named the **Distinguishing Approach**, the **Indexing Approach**, and the **Information Structure Approach**.

### 2.2 Transitivity

The domain of transitivity has been intensely investigated since Antiquity. Traditionally, the distinction between transitivity and intransitivity has been traditionally viewed as a difference in valence. Transitive verbs must have two arguments, an agent and a direct object, as opposed to intransitive verbs, which require only one argument to form a grammatical sentence. This structural view has been complemented by a semantic definition of transitivity. In line with the etymon of the word (Lat. *transire*: "pass over"), a transitive verb has been semantically characterised as an action whose effects pass over from an entity that does something (in my terminology, an agent) to an entity who is somehow affected by the action carried out by the agent.

Various criteria have been proposed in order to account for the difference between transitive and intransitive verbs. Many of these criteria are structurally-based, such as the possibility for clauses to be passivised, or the possibility for direct objects to take accusative case. However, as noted by Kittilä (2002: 32-33), the use of structural criteria for defining transitivity is highly problematic both from an intra-linguistic and a cross-linguistic point of view. Languages indeed exhibit a large amount of variation as to, e.g., passivisation. Many languages lack passive constructions or may apply passivisation even to intransitive clauses, giving thus impersonal passives (see Siewierska 1984, 2010). Moreover, criteria such as the possibility to take accusative case do not cover many cases attested across languages. For example, as noted by Kittilä (2002: 77) the second participant of a transitive clause can be encoded in many different ways: it can be coded with a case other than the accusative (as the use of dative-coded objects with some verbs in German, 2.1), or it can be coded by an adposition, as in English (2.2):

(2.1) German (Indo-European, Germanic)

(a) *Er schläg-t den junge-n*  
 he.NOM hit.PRS-3SG DET.ACC boy-ACC  
 "He is hitting the boy"

(b) *Er hilf-t der frau*  
 NOM help.PRS-3SG DET.DAT woman  
 "He is helping the woman" (Kittilä 2002:77)

(2.2) English (Indo-European, Germanic)

- (a) He killed the boy.  
 (b) He looks at the boy.

While the verb *schlagen* in German takes an accusative direct object, the verb *helfen* requires its direct object to be dative case-marked. Similarly, the verb *kill* in English is followed by a bare direct object, whereas the verb *look* requires its direct object to be preceded by the preposition *at*. However, this characterisation largely rests upon the concepts of "subject" and "direct object", whose theoretical and cross-linguistic status is highly controversial (Croft 2001, Cristofaro 2009, 2010, Dryer 1997), as I will discuss in the next section. A semantic approach to transitivity has proved to be more fruitful, especially in dealing with the tremendous diversity found within the 'transitive' domain across world's languages. The best-known characterisation of semantic transitivity is that proposed by Hopper

and Thompson (1980) <sup>1</sup>. Based on a study of transitivity in a number of languages, they view transitivity as a multifactorial, scalar phenomenon, determined by the interplay of parameters listed in table (2.2).

	<i>high</i>	<i>low</i>
A. Participants	2 or more participants	1 participant
B. Kinesis	action	non-action
C. Aspect	telic	non-telic
D. Punctuality	punctual	non-punctual
E. Volitionality	volitional	non-volitional
F. Affirmation	affirmative	negative
G. Mode	realis	irrealis
H. Agency	A high in potency	A low in potency
I. Affectedness of O	O totally affected	O non affected
J. Individuation of O	O highly individuated	O non individuated

Table 2.2: Transitivity parameters (Hopper and Thompson 1980: 252)

INDIVIDUATED	NON-INDIVIDUATED
proper	common
human, animate	inanimate
concrete	abstract
singular	plural
count	mass
referential, definite	non-referential

Table 2.4: The Individuation Hierarchy (Timberlake 1977: 162)

<sup>1</sup>In fact, Lakoff (1977: 244) had proposed a model similar to that proposed three years later by Hopper and Thompson. He provides a *Gestalt* model of transitivity as a cluster of semantic properties typical of transitive constructions: 1. there is an agent, who does something 2. there is a patient, who undergoes a change to a new state (the new state is typically unexpected) 3. the change in the patient results from the action by the agent 4. the agent's action is volitional 5. the agent is in control of what he does 6. the agent is primarily responsible for what happens (his action and the resulting state) 7. the agent is the energy source in the action; the patient is the energy goal (that is, the agent is directing his energies towards the patient) 8. there is a single event (there is spatio-temporal overlap between the agent's action and the patient's change) 9. there is a single, definite agent 10. there is a single, definite patient 11. the agent uses his hands, body, or some instrument 12. the change in the patient is perceptible 13. the agent perceives the change 14. the agent is looking at the patient.

In their approach, transitivity is a gradient phenomenon, influenced by interplay of semantic, pragmatic and syntactic properties. A clause can be more or less transitive based on the features it shows with respect to the parameters listed above. A clause that shows all the parameters in the high column is the most representative example of a transitive clause: a prototypical transitive clause is thus characterised not only as the complete transfer of an action from an agent to a patient, highly distinct from each other. It is further determined by the presence of other features as well, such as telicity, punctuality, and reality of the action. The agent will be highly individuated (i.e. animate and definite, see below) and volitional, since she has control of the action (see table 2.4). The patient, on the contrary, will be highly affected by the action performed by the agent: in their view, the more individuated a patient is, the more affected by the action it will tend to be. Therefore, sentences with verbs like "kill, break, move" will be considered as more transitive than sentences with predicates like "hit, search, love", since the former involve a change of state or position in their direct object participant, whereas the latter either do not involve a change (as with "love") or do not specify if the direct object has undergone some change (as with "hit", see Levin and Rappaport Hovav 2005). The semantic difference between more and less transitive predicates is reflected in morphosyntax. Direct objects governed by less transitive verbs indeed tend to take cases other than accusative or absolutive, or to be coded as an oblique or incorporated. The examples (2.1) from German discussed above clearly illustrate this pattern.

Hopper and Thompson thus see as a property of argument structure, rather than a relationship of the verb with its direct object. The most important aspect of their characterisation of transitivity lies in the prediction they make about the relationship between the features. They claim that component features should co-vary extensively and systematically, so that "whenever two values of the transitivity components are necessarily present [...] they will agree in being either both high or both low in value" (Hopper and Thompson 1980: 254). They therefore propose the following Transitivity Hypothesis:

"If two clauses (a) and (b) in a language differ in that (a) is higher in Transitivity according to any of the features A-J, then if a concomitant grammatical or semantic difference appears elsewhere in the clause, that difference will also show (a) to be higher in Transitivity" (Hopper and Thompson 1980: 255).

Hopper and Thompson (1980: 280 ff.) further claim that degree of transitivity of a clause affects discourse organisation. They claim that high transitivity tends to correlate with foreground, whereas low transitivity correlates with backgrounding. The presence of two participants is thus claimed to be a strong indicator of foregrounding, as opposed to the presence of one participant, which instead

indicates backgrounding. For instance, they show that individuated (that is, definite, animate or referential) and affected direct objects are strongly correlated with high transitivity and therefore with foregrounding. This semantic and discourse prominence is likely to be formally coded in many languages. DOM is one of the phenomena that they use to illustrate the connection between semantic and formal transitivity. Consider the following examples from Spanish:

(2.3) Spanish (Indo-European, Romance)

(a) *Celia quiere mirar un bailarín*  
 Celia want.PRS.3SG watch.INF a dancer  
 "Celia wants to watch a dancer"

(b) *Celia quiere mirar a un bailarín*  
 Celia want.PRS.3SG watch.INF DOM a dancer  
 "Celia wants to watch a (particular) dancer" (Hopper and Thompson 1980: 256)

(2.3a) and (2.3b) diverge with respect to their morpho-syntactic coding, since the direct object in (2.3b) but not that in (2.3a), is introduced by the preposition *a*. This formal difference reflects a semantic difference in definiteness. In (8.45a) the direct object is indefinite and not-referential, while in (8.45b) the direct object is referential and therefore more individuated. The presence of the preposition indicates that the speaker presumably knows the identity of the referent. According to the Transitivity Hypothesis, (2.3b) is higher in transitivity than (2.3a), since a more individuated direct object is taken as a feature of high transitivity and foregrounding.

Although Hopper and Thompson's work on transitivity paved the way for many studies focussing on verbal semantics, argument structure, and case marking phenomena, nonetheless some points of their Transitivity Hypothesis have been strongly criticised (see Kittilä 2002 for a useful summary). Here I will discuss only the aspects relevant to the examination of DOM. One major criticism of Hopper and Thompson's work is that it fails to provide a clear division between transitive and intransitive clauses. They indeed state that an intransitive clause like "Susan left" can be regarded as more transitive than a transitive clause like "Jerry likes the beer" (Hopper and Thompson 1980: 254), since the intransitive clause displays four transitivity components, namely kinesis, telicity, activity, and volitionality. On the contrary, the only transitive feature of the transitive sentence is the presence of two participants. However, as noted by Lazard (2002) and Kittilä (2002: 118), among others, the presence of two participants is a fundamental attribute of transitive clauses as opposed to intransitive ones, both from a syntactic and semantic perspective. Therefore, only situations involving at least two

participants ought to be considered when investigating transitivity and related phenomena. Another *punctum dolens* of their approach is the alleged co-variation between features. Lazard (1994), following a previous proposal by Tsunoda (1985), challenges the view that all parameters are equally important in determining the morpho-syntactic expression of transitivity, and proposes that the most relevant parameters for investigating the coding of transitivity cross-linguistically are

1. the presence of two participants distinct from each other;
2. the agentivity and volitionality of the agent on the event described by the predicate;<sup>2</sup>

Finally, there has been an intense debate about the alleged co-variation between features proposed by Hopper and Thompson (1980). Whereas it could be quite intuitive that features (E) and (J) co-vary, in that they pertain to the agent, it is far less clear that volitionality and individuation of the direct object co-vary, since these two parameters are related to different participants of the clause. Malchukov (2006) therefore suggested a revision of Hopper and Thompson's original hypothesis. In his model, represented in table (2.6), parameters are grouped around the element of the clause they refer to, so as to make clear "the mutual semantic affinities between particular semantic parameters" (Malchukov 2006: 333).

A-parameters	V-parameters	O-parameters
[animacy][volitionality]	[kinesis][factivity][tense/aspect]	[affectedness][individuation]

Table 2.6: Transitivity scale (Malchukov 2006: 333)

Thus, there are parameters related to the Agent, parameters related to the verb, and parameters related to the direct object. However, as rightly observed by de Swart (2007: 31), Malchukov's revision does not take into consideration that, in Hopper and Thompson's model, animacy is a feature related not only to the agent but to the direct object as well, insofar as animacy is one of the essential properties of individuation, along with definiteness and referentiality.

The semantic characterisation of the participants of the prototypical transitive construction has been the object of an intense debate within linguistics during the last two decades. As I will discuss in the next sections, many studies challenge the idea that the prototypical direct object should be as animate and definite as

<sup>2</sup>See Tsunoda (1985) for a critical discussion about the role of the agentivity and volitionality parameters in determining transitivity.

3. the affectedness of the patient



the agent. According to Comrie (1989: 128) and Langacker (1991: 283), a prototypical transitive clause has many of the properties identified by Hopper and Thompson (1980), in that it describes an event in which an agent (Langacker's *energy source*) causes a patient (Langacker's *energy sink*) to undergo a change of state or place. Both participants involved in the event have to be discrete entities, highly distinguishable from each other. As for their semantic properties, the prototypical agent should be volitional, which, in turn, presupposes animacy (or, rather, humanness). Crucially, Comrie's and Langacker's characterisation of the prototypical direct object differs from Hopper and Thompson's in that, in their view, the prototypical direct object is inanimate, or at least less animate and definite than the agent. This view derives from the different roles that participants play in the prototypical transitive event. While the prototypical agent must be human or at least animate in order to volitionally initiate an action, the prototypical patient does not take part actively in the event, aside from receiving the effects of the action and thus undergoing some visible change of state or position. The claim that direct objects are usually less animate and definite than agents could therefore result from the asymmetry between the two participants. In these terms, this inactive vs. active asymmetry is reflected by the semantic properties of the two participants. The human, or animate, participant, will be encoded as the agent, whereas the less animate or definite participant will be encoded as the direct object. Deviations from this situation lead, according to Comrie (1989), to differences in i) the linguistic coding of the core participants of a clause (case-marking and indexation strategies), ii) verb morphology, and iii) word order.

The identification of the semantic and pragmatic properties of prototypical direct objects has been crucial to the study of DOM, since explanations of DOM heavily rely on which features prototypically correlate with direct object. As will be discussed more thoroughly in the next sections, two main perspectives have, in fact, been assumed in linguistic theory concerning DOM. On the one hand, DOM has been taken as an indicator of high transitivity, since highly individuated direct objects are central for defining transitivity, as advocated, for instance, by Hopper and Thompson (1980). On the other hand, DOM has been taken as a reflex of the markedness of animate and definite direct objects, since prototypical direct objects are characterised as inanimate and indefinite (Comrie 1989, Croft 1988). The "high transitivity" approach stems from the so-called "indexing" function of case marking on core arguments, whereby case marking indexes semantic or pragmatic properties of the referent of arguments, such as animacy, definiteness, and topicality (Siewierska and Bakker 2008: 292). The "markedness" approach assumes that the basic function of case marking on core arguments is to differentiate agents from direct objects. In this view, only one of the two core arguments in transitive clauses receives case marking, while the other argument remains uncoded (Comrie 1989, Dixon 1994, among others). The two analyses of DOM thus

seem to be in direct conflict with each other, since they argue for very different characterisations of direct objects and for different functions of DOM.

Before discussing the previous approaches to DOM in detail, it is worthwhile discussing some issues relevant to the characterisation of the grammatical relation of direct object.

### 2.3 The puzzle of direct object

Over the last decades, grammatical relations (henceforth, GRs) have been the object of several studies both from functional and formal perspectives (Blake 1994: 48-93, Bresnan 2001: 44-60, Farrell 2005, Fillmore 1968, Givón 1984, 1997; 2001: 173-232; Keenan 1976, Perlmutter 1983, among others).

Many linguists of different theoretical persuasions have attempted to provide cross-linguistic definitions of grammatical relations based on either a set of formal criteria, such as indexation or raising, or semantic concepts. The basic idea underlying these approaches is that if different languages have a grammatical relation that shows certain selected properties, then these languages can be said to show the same grammatical relation. Therefore, GRs are universal categories whose cross-linguistic existence is taken for granted, without paying much attention to the non-overlapping properties that GRs may display (Cristofaro 2009). In formal approaches, such as Generative Grammar, this idea has been primarily developed on a theory-internal basis, for example postulating that the subject is the NP occupying the specifier position under the IP node. Conversely, in functional approaches, GRs have often been regarded as prototypical categories, i.e. a cluster of properties which are not obligatorily shared by all of the member of the category. Both in formal and functional approaches, the criteria for identifying GRs (as well as other linguistic categories) are often based on distributional analyses, whereby a category is identified based on its occurrence in different constructions, provided that its distribution is the same with regard to the possibility vs. impossibility to occur in a context. For instance, some standard criteria for the identification of objecthood are passivisability, overt expression in transitive clauses, case marking, the position with respect to the verb, and pronominalisation, as shown by the examples in (2.4):

(2.4) English (Indo-European, Romance)

- (a) He killed the boy
- (b) The boy has been killed (passivisation)
- (c) \*He killed
- (d) He killed the boy (overt expression in transitive clauses)

- (e) \*She loves he
- (f) She loves him (accusative case of the pronoun as opposed to nominative *he*)
- (g) \*Him loves she
- (h) She loves him (postverbal position of the DO)

As noted by Croft (2003), and Cristofaro (2009), among others, these criteria offer inconclusive or partial evidence for direct objecthood when used to draw cross-linguistic generalisations. Moreover, they could give very inconsistent results when used to determine the direct object status even in only one language. As is well known, measure phrases, which in many languages are syntactically encoded as DOs, cannot undergo passivisation, as in (2.5):

(2.5) English (Indo-European, Romance)

- (a) I weigh 80 kilos
- (b) \*80 kilos are weighed by me

The same problem arises when we use case marking as a means to identify the grammatical relation of direct object. Direct objects are usually distinguished from other GRs based on the fact that they trigger the use of accusative case or indexation (in languages with case systems or indexation of direct objects, of course). However, a noun in direct object position may be coded by another case, either because the verb requires the direct object to be in another case (as we have seen in the case of the German verb “*hilfen*” as in 2.1<sup>3</sup> or because the use of different case marking indicates a semantic contrast, as in the following example from Icelandic:

(2.6) Icelandic (Indo-European, Germanic)

- (a) *Hann klóraði mig*  
he.NOM scratched me.ACC  
"He scratched me"
- (b) *Hann klóraði me*  
he.NOM scratched me.DAT  
"He scratched me" (Barddal 2001)

The contrast between (2.6a) and (2.6b) has been interpreted in terms of affectedness vs. control of the action. When coded by the accusative, the direct object is completely affected by the action, and the scratching is construed as

<sup>3</sup>Note that while “*hilfen*” can be passivised, the direct object still maintains its dative coding.

painful and violent. However, the use of dative indicates that the patient has some control over the action, that is seemingly 'in control of its own affectedness' (Næss 2004: 1205). The problem of identifying GRs is even trickier when dealing with possible cross-linguistic characterisations, since often such definitions are based on a small number of selected properties and disregard differences and non-overlapping properties. For instance, direct-inverse languages challenges the common criteria used to define GRs, since their morphological encoding is governed by the ranking of agents and direct objects with respect to the position of each on the animacy hierarchy (see Zúñiga 2006).

One viable solution to this problem is to assume that GRs, as well as linguistic categories, are language-specific, in the sense that "particular grammatical relations, e.g. subject, can only be posited for individual languages, and the notion of grammatical relation cannot be cross-linguistically valid either in the sense that all languages have the same grammatical relations, or in the sense that grammatical relations found in different languages should be expected to be instances of the same grammatical relation" (Cristofaro 2009). In this approach, first advocated by Dryer (1996, 1997) and later developed by Croft (2001), Cristofaro (2009, 2010), and Haspelmath (2007, 2010), GRs such as direct object are no more than convenient descriptive labels signalling that some linguistic elements display some selected properties cross-linguistically. Hence, there is no need to postulate the existence of cross-linguistic categories, nor are they part of a speaker's linguistic knowledge (see Cristofaro 2009 for a detailed discussion of the alleged mental reality of linguistic categories). As demonstrated by Dryer (1997: 123), this approach is fully compatible with a functional-typological analysis, since the similarities displayed by linguistic categories across different languages are ultimately due to functional and cognitive principles that are valid cross-linguistically. The use of labels such as "subject" and "direct object" can be maintained as long as this use does not obscure the fact that we are dealing with different and language-specific categories, without any assumption of their cross-linguistic status. A somehow more radical approach has been advocated by proponents of Construction Grammar. Besides being language-specific, GRs can be construction-specific as well, in that different constructions (i.e. pairings of forms and function, see Croft 2001) in the same language define different GRs.

This work fully espouses the language specific view of GRs. The grammatical relation of DO will thus be defined in terms of some semantic and pragmatic properties that are shared by all of its members. Yet, I believe that formal, i.e. morpho-syntactic, criteria are crucial for the cross-linguistic comparison of any linguistic phenomenon. I will try to provide a general characterisation of direct object properties, with particular regard to the semantic and pragmatic properties.

## 2.4 Approaches to direct objecthood

Only few studies have been devoted to the properties of the grammatical relation of direct object, (see Plank 1984, Lazard 1994, 1998, 2003), as opposed to those devoted to the notion of subject (cf., for instance, Li 1976, Keenan 1976, Farrell 2005 and references therein, among others). Since the main aim of this work is to investigate how semantic and pragmatic properties of direct object referents can affect the morphosyntax of DOM, and how languages are more or less sensitive to such features, here I will focus on the discussion found in the literature concerning these properties. As we have seen in section 2.2, the direct object is prototypically a patient or a theme, that is the participant that undergoes a change of state or location. Hence, the grammatical relation of direct object is closely related to the semantic role of patient. Indeed, every characterisation of the grammatical relation of direct object links objecthood to patienthood and themehood (Dowty 1991, Fillmore 1968, van Valin and LaPolla 1997, Farrell 2005, Primus 1999). Most linguists would agree on the characterisation of direct objects as affected participants.

A fruitful and detailed definition of direct object has been proposed by Croft (to appear). Croft's model of event structure and argument realisation is primarily based on the notions of force dynamic relations among participants and of verbal profile (i.e. denoted by the verb, see Talmy (1976, 1988); Croft 1991, to appear). In his model, a direct object is defined as the endpoint of a linear causal chain (i.e. 'the sequence of participants linked by force dynamic relations that hold among them) characterised by the transmission of force (force dynamic relations) from one participant (the initiator) to another (the endpoint) as a directed change (Croft to appear: 276). The dimension of directed change is crucial in defining direct objecthood and in predicting the realisation of a participant as a direct object, as it concerns the direction of change from inception to completion and encompasses various parameters such as telicity, punctuality, and durativity (Croft to appear, ch. 3, pp. 1 ff.; Croft 2010: 10 ff.). Croft states that arguments are more likely to be realised as direct objects whenever:

The prototypical simple verb lexicalization would possess a single completed directed change that is the endpoint of the transmission of force. Deviation from the prototype leads to a greater likelihood of realization of the participant as a (Subsequent) Oblique rather than an Object, although Object realization remains possible for a participant lacking all of these properties. (Croft to appear), ch. 8).

Therefore, a prototypical direct object can be defined as the affected participant of a telic, punctual event involving some transmission of force. As we will see in the next chapter, Croft's definition of direct object as the endpoint of an event

involving transmission of force will be particularly useful for explaining the diachrony of object markers. Another important aspect of Croft's model is the fact that the mapping from event structure (i.e. the causal chain and verbal profile) to argument roles is direct, and therefore there is no need to posit any semantic role. This approach has the advantage of avoiding the imposition of further levels of analysis. Cross-linguistic and intra-linguistic variation in the coding of the direct object (especially as concerns object incorporation and oblique encoding of direct objects) can thus be regarded as the reflection of the relative position in the causal chain, as well as the verbal profile of the event. (Croft to appear: 180). Furthermore, the definition provided by Croft fully aligns with a language-specific approach to grammatical categories, since this definition is based on universal conceptual-semantic and general notions.

Before turning to the discussion of the semantic and pragmatic properties that are most often associated with the grammatical relation of direct object, it is worth mentioning another important distinction concerning object roles proposed by Dryer (1986) in a seminal paper. Through the investigation of the expression of direct objects in transitive clauses and recipients and themes in ditransitive clauses, Dryer identifies two strategies that are used across languages to code these argument roles. The first strategy involves the opposition between the grammatical relations of direct object and indirect objects (DO/IO opposition), and is commonly found in European languages. the theme of a ditransitive clause (T) is coded in the same way as the patient of a transitive clause (P), by coding both as accusative (or absolutive), for example. The recipient is instead coded differently (R) by, e.g., dative case. Other languages group these three roles differently: recipients of ditransitive clauses and patients of transitive clauses are coded in the same way, denoted as primary object (PO), while the theme, called a secondary object (SO), receives a different coding.

The explanation provided by Dryer for this variation of the coding of object roles is based on grammaticalisation processes. The DO/IO strategy derives from the grammaticalisation of the semantic role of the arguments (since both the patient and the theme are the participants that undergo a change of state or location). In contrast, the PO/SO pattern is motivated by their topicality: the primary object encompasses the roles which are usually higher on the topicality hierarchy, while the secondary object is either less topical or not topical at all. As I will argue in the next chapters, the PO/SO and DO/IO distinction seems to influence the use of DOM in ditransitive clauses.

Let us now turn to discuss some proposals that have been made about the semantic and pragmatic properties of direct objects. As we have seen at the end of section 2.2, the identification of the semantic and pragmatic properties that are most often associated with the grammatical relation of direct object has been

heavily debated in the literature, and the various attempts to define direct object properties are in conflict with each other.

Givón (1979) hypothesised that case-marking systems serves to encode the syntactic function as well as the semantic and pragmatic role of the NPs required by verb valency. In his view, both “subject” (i.e. agent) and “direct object” are grammaticalised categories related to topicality effects. The choice of a particular argument as direct object (or as a subject/agent) is based on the notion of “topic continuity”. The degree of topic continuity is measured in terms of referential distance between the previous instances of an NP and its current occurrence, the persistence of a NP in subsequent discourse, and its referential ambiguity, (the number of competitors for topic positions in the clause). Givón employs the notion of topic continuity to account for the selection of subject and direct object, as well as to describe word order alternations and voice changes. In his model, the more topical NP will be assigned the role of agent (primary topic), whereas the less topical NP will be selected as a direct object (secondary topic) (Givón 1984: 135). Givón exemplifies the role of topic continuity in determining direct object choice with the so-called “dative-shift” construction, that is the phenomenon whereby an oblique argument functioning as a recipient or benefactive (usually coded by the dative case) is encoded as a direct object, as in example (2.7):

- (2.7) a. *Context*: Who did Mary give the book to? (PAT-topic, DAT-focus)  
 b. *Reply*: She gave the book to Bill (PAT is DO)  
 c. ?She gave Bill the book (?DAT is DO)  
 d. *Context*: What did Mary give to Bill? (DAT-topic, PAT-focus)  
 e. *Reply*: She gave him a book (DAT is DO)  
 f. ?She gave a book to him (?PAT is DO)

In (2.7), the choice of coding the animate argument as a direct or indirect object is, according to Givón, directly related to the degree of topicality of the argument. In (2.7b) the direct object is the (secondary) topic of the question and the dative in focus. Dative shifting thus sounds inappropriate as a reply to (2.7a). Conversely, in (2.7e), the recipient/dative is topical and the patient/theme is in focus. Therefore, the recipient/dative is coded as direct object by being moved to the direct object position. In Givón’s view, the selection of participants as subject, direct object or oblique is directly governed by topicality properties. The more topical NP will be selected as the agent, and less topical NPs will be selected, respectively, as direct objects or obliques. The grammatical relation hierarchy subject<direct object<oblique (initially posited to account for patterns of relativisation across languages under the name of “NP-Accessibility Hierarchy”, (see Keenan and Comrie 1977) can thus be reinterpreted as a hierarchy of topicality. Nonetheless, Givón (1976, 1991) proposed a different version of the topicality hierarchy:

## (2.8) agent &gt; dative/benefactive &gt; accusative/patient

This hierarchy is motivated by the fact that datives and benefactives tend to be expressed as uncoded direct objects in languages where dative shifting or mechanisms of promotion to direct object position is available. However, as noted by Givón (1979) and Croft (1991: 153), the more topical status of datives/benefactives as opposed to patients/direct objects seems to be due to the fact that datives are overwhelmingly human or animate.

It should be noted, however, that direct object position is exploited to introduce new participants into discourse, thus showing a correlation with the focus position of the clause (Dubois 1987). Indeed, as will be shown in the next chapters, there seems to be a striking correlation between topical direct objects and the presence of DOM on the one hand, and between focal direct objects and the absence of DOM on the other one.

Before turning to the issue of the semantic properties of “prototypical” direct objects, which is the nucleus around which the various approaches to DOM have been revolving, I would like to briefly discuss Lazard’s proposal about a viable description of direct objects from a cross-linguistic perspective (Lazard 1994, 2001, 2003, 2006).

Since it is impossible to provide a cross-linguistically valid definition of direct object based only on morphosyntactic criteria, Lazard proposes to use an “arbitrary conceptual framework” as a tool to deal with cross-linguistic variation (Lazard 2001). The arbitrary conceptual framework used to investigate transitivity, and the expression of direct objects, is what Lazard calls the “prototypical action”, i.e. an action performed volitionally by a human agent which causes a change in an individuated patient. In Lazard’s view, direct objects are best characterised in terms of a radial category, an “object zone”. The direct object of a prototypical action will always be referential and “autonomous” both syntactically and semantically, in that it can be moved to virtually every position within the clause. This is what Lazard (2003: 13) calls a “distant object”. Non-referential direct objects instead exhibit a strong tendency to be very close to the predicate, i.e. it is not possible to move them freely within the clause, and they are usually very low in individuation, what Lazard calls “close objects”. In a sense, such direct objects are seen as strongly depending on the predicate, as opposed to distant objects, whose animacy and referentiality make them more prominent. Lazard exemplifies this distinction with the example (2.9) from Persian, in which there are two objects: one is uncoded, while the other one can be alternatively overtly coded or uncoded. In (2.9), the first direct object, “*pul-eš-ra*” is referential and autonomous with respect to the verb, and is therefore overtly coded by postposition *-ra*. The second object “*širini*” is uncoded and cannot be moved, since it is low



in referentiality and a sort of "qualifier" of the verb (Lazard 1994), thus closely resembling the phenomenon of incorporation:

(2.9) Persian (Indo-European, Iranian)

*pul-eš-ra            širini    mixor-ad*  
 money-his-DOM sweets eat:PRS-3SG

"He spends his money in eating sweets" (lit. He eats sweets his money;  
 Lazard 2003: 4)

In Lazard's model as well, direct objects are defined in terms of high individuation and change of state. But, as we will see shortly, considering highly individuated direct objects as the 'prototypical' direct objects is in direct conflict with Comrie's and Croft's characterisation of direct object on the one hand, and with markedness theory on the other hand. Highly individuated direct objects are the ones which receive overt morphological coding in DOM, and morphological overt coding is taken, in markedness theory, as the reflection of a marked feature of these direct objects.

## 2.5 Differential object marking: previous approaches

In this section, previous approaches to DOM will be presented and compared, and some of the main problems and inconsistencies will be discussed. Three main approaches to DOM can be identified: the **Distinguishing** Approach, (which takes DOM as a means of differentiating between arguments of a transitive clause); the **Indexing** Approach, (according to which case marking signals properties of the referents of arguments or of the clause, such as animacy or definiteness); and, the **Information Structure** Approach, (which takes DOM as a reflex of the secondary topic status of overtly coded direct objects, see below). These approaches (except for the information structure approach, which can be traced back to either of the two other approaches) stem from the general explanations that have been proposed in order to account for the function of case marking and indexation phenomena, as briefly discussed in section 2.2.

### 2.5.1 The Distinguishing Approach

The view that case marking of core arguments has a distinguishing function has a long history in linguistics (see, for instance, Diez 1863, Meyer-Lübke 1900). The basic idea underlying this approach is that case marking primarily serves to avoid ambiguity in the assignment of grammatical relations within the clause

(Comrie 1989, Dixon 1979, 1994, among others). That is, case marking is used on core arguments of transitive clauses in order to allow the hearer to distinguish the subject/agent from the direct object/patient. In this view, the use of case marking is minimal, as it only serves to differentiate between agents and direct objects. Usually only one of the two core arguments in a transitive clause receives case marking, while the other argument remains uncoded. It is also possible for both of the arguments to receive different case markings, such as in the Latin declension case system. By contrast, overtly coding the only argument of an intransitive predicate (i.e. the intransitive subject) is unnecessary, as such an argument cannot be confused with other arguments in the clause. The most common situation cross-linguistically, however, is for both arguments to be zero-coded, as shown by the sample in the WALS, in which 98 languages show neutral alignment (Comrie 2008, ch. 98).

DOM systems lend themselves to be analysed within this approach, since there could be a need to disambiguate when both the subject and the direct object are animate and/or definite. When both agent and direct objects share the same semantic properties, and other clues for identifying the NP's syntactic role are not found in the clause, additional strategies are required in order to allow for the correct interpretation of grammatical relations. Under this view, there would be no difficulty in assigning the correct grammatical relations in sentences like *Mary broke the bottle*, as the agent is the animate NP whereas the direct object is the inanimate and affected NP. Since participants in the event described by the clause are consistent in their semantic and pragmatic properties, no overt coding is usually required to identify the grammatical relations in the clause. World knowledge helps the hearer assign the correct interpretation. Problems arise, however, when the NPs with similar properties are found within the same clause, namely when the properties of the direct object resemble those of the agent, such as in *Richard beat up Catherine*, where both arguments have high animacy and definiteness. In such cases, overt coding is likely to be used for direct objects in languages that have a DOM-system, since these direct objects exhibit properties usually associated with agents. It should be recalled that, according to the mainstream functional literature, prototypical transitive constructions are assumed to consist of a volitional agent -usually animate and definite- and a direct object which is fully affected by the action carried out by the agent, while being less animate and definite than the agent<sup>4</sup> ((e.g. Bossong 1985, Comrie 1979, Croft 1988, among others). Therefore, within this approach, DOM has been assumed to reflect the marked status of definite and animate direct objects as opposed to indefinite and inanimate ones.

---

<sup>4</sup>Cf. Comrie (1989: 128): “The most natural kind of transitive construction is one where the A is high in animacy and definiteness, and the P is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction [...] The construction which is more marked in terms of information flow should also be more marked formally’.

The approach discussed so far originates in the typological markedness theory first described by Greenberg (1966) and later elaborated by Croft (1990, 2003), among others, in which marked forms express marked meanings (see chapter 1 for a brief discussion of the concept of markedness). As far as DOM is concerned, this means that inanimate and indefinite objects are *conceptually* unmarked with respect to transitive events, insofar as they are easily distinguishable from the agents. The presence of an additional morphological marker on direct objects that are high in individuation iconically signals the semantically marked status of these objects with respect to objects low in individuation. This type of case marking is thus economically motivated, as no overt coding is required when there is no need to disambiguate between the NPs.

Let us consider the case of Yongren Lolo, a Tibeto-Burman language spoken in China (Gerner 2008). Yongren Lolo is a verb-final language with otherwise very free word order (Gerner 2008: 298). According to the author, DOM, expressed by the particle  $t^hie^{21}$  postposed to the direct object, is used only when the predicate frame is ambiguous as to the assignment of the core argument roles. Indeed, Gerner (2008: 299) lists three main contexts in which DOM is likely to appear:

- Ambiguous predicational frames are disambiguated by the particle  $t^hie^{21}$  postposed to O;
- The particle  $t^hie^{21}$  appears after the O of an ambiguous frame even if the A is a zero-anaphora NP;
- The word order of A and O- $t^hie^{21}$  is free before the verb.

Consider the examples in (2.10):

(2.10) Yongren Lolo (Sino-Tibetan, Tibeto-Burman)

(a)  $\eta o^{33}$   $\epsilon \epsilon^{33} mo^{33}$   $t^hie^{21}$   $t^s^h o^{33}$   $zi^{33}$   
 1P.SG snake DOM follow go

"I will follow the snake"

(b)  $\eta o^{33}$   $\epsilon \epsilon^{33} mo^{33}$   $t^s^h \text{ } \text{ }^{33}$   $zi^{33}$   
 1P.SG snake follow go

"The snake follows me" (Gerner 2008: 299)

In Gerner's analysis, the use of the object marker  $t^hie^{21}$  in (2.10a) serves to correctly allocate the roles of agent and direct object to "I" and to "snake" respectively. As pointed out by Gerner (2008: 299), the sentence would be perfectly grammatical even without the object marker, as in (2.10b), but in this case ambiguity would arise, since an interpretation like 'The snake follows me' is acceptable

based on world knowledge, as shown by the difference between the two examples. Moreover, since Yongren Lolo has a free word order, the ordering of arguments does not offer any clues about the interpretation. The contrast between (2.10a) and (2.10b) clearly shows that the object marker has a distinguishing function, since it ensures one interpretation over the other. In Yongren Lolo, DOM is thus used to facilitate the interpretation of sentences, whenever the arguments display the same semantic features.

The distinguishing function of case marking is closely interwoven with the animacy and definiteness of direct objects. As I will discuss below, these properties are usually represented through a hierarchy, originally proposed by Silverstein (1976), which actually conflates three different hierarchies, namely person, animacy, and definiteness (Croft 2003: 130):

- (2.11) EXTENDED ANIMACY HIERARCHY: first/second person pronoun > third person pronoun > proper names > human common noun > non-human animate common noun > inanimate common noun

According to the view advocated by Comrie (1989) and Dixon (1994), among others, nominals on the left-hand of the hierarchy are more likely to be selected as agents, since they are high in both animacy and definiteness, whereas nominals on the right-hand are more likely to be chosen as direct objects, since they are low in both animacy and definiteness.

This view is supported by the structural markedness criterion, as highly animate and definite direct objects receive overt coding as opposed to indefinite and inanimate ones. The frequency criterion also provides further support to the marked status of animate and definite direct objects. Filimonova (2005: 78) quotes Thompson (1909) on statistics about the use of animate and inanimate nouns as agents and direct objects in a corpus of Russian provided by Thompson (1909). Thompson found that only 10% of direct object referents are animate, as opposed to 75% of agents. Based on these findings, he proposed a hierarchy that anticipated later formulations found in linguistic typology:

- Agent ←— persons children animals inanimates —→ Patient

This hierarchy, like the extended animacy hierarchy above, predicts that the more animate nominals are likely to occur as agents, while inanimate referents are more likely to occur as direct objects. A situation contrary to this pattern should be overtly coded to disambiguate (see Filimonova 2005 for a discussion). More recently, the extensive use of corpora has made it possible to investigate the textual frequency of animate and definite direct objects. For instance, Dahl's (2007: 118) counts of a Swedish corpus show that direct object referents are overwhelmingly

inanimate (89%), as opposed to agent referents, which are overwhelmingly animate (92%). Direct objects seem to be closely correlated with a low degree of animacy and definiteness with regard to markedness criteria<sup>5</sup>. The direct object in (2.12a) should thus be considered as semantically and cognitively less marked compared to the one in (2.12b):

(2.12) Spanish (Indo-European, Romance)

- (a) *Vio*            *el* *perro*  
 see:PST.3SG the dog  
 "He saw the dog"
- (b) *Vio*            *a*    *ti*  
 see:PST.3SG DOM 2SG  
 "He saw you"

However, this approach carries with it various well-known limitations. First of all, as noted by Moravcsik (1978), languages can tolerate some amount of ambiguity between subject and direct object, since it is often possible to assign an NP to a given grammatical relation based on world knowledge or context<sup>6</sup>. Moreover, some constructions are hardly analysable in terms of a strict distinguishing function of case marking. Let us briefly consider the well known case of noun incorporation. It has been noted that direct objects low in individuation are more likely to be incorporated (Mithun 1984, Hopper and Thompson 1980). Clauses with incorporated direct objects are often intransitive from a syntactic point of view, even though they depict an event with two participants<sup>7</sup>. For instance, in Selayarese (Austronesian, Finer 1997), *definite* direct objects are indexed on the verb, as in (2.13a). When the direct object is indefinite, there is indexation only with subject and the verb takes the intransitive prefix *ang/a?* as in (2.13b; see section 6.11):

(2.13) Selayarese (Austronesian, Sulawesi)

- (a) *u-alle-i*      *doe?-ijo*  
 1SG-take-3 money-the  
 "I took the money"

<sup>5</sup>Cf. Croft 1988: "The natural correlation of direct object is with low animacy, low definiteness, and highly affected objects, i.e. genuine patients".

<sup>6</sup>Cf. Moravcsik (1978: 255) "The precise threshold of subject-object ambiguity tolerance in natural language is unknown at present".

<sup>7</sup>Logan Sutton brought to my attention that Kiowa-Tanoan languages, for instance, do not decrease the valence of a verb when a noun is incorporated (at least as far as pronominal indexation on the verb goes).

- (b) *(a)ng-alle-kang doe?*  
 INTR-take-1PL money  
 "We took some money" (Finer 1997: 679-680)

This pattern seems to be quite common cross-linguistically. Many languages resort to detransitivising constructions when the direct object is indefinite. As noted by Næss (2004: 1192), the idea that direct objects low in individuation are the most typical ones could be highly problematic, since these clauses do not involve a direct object at all from a structural point of view. In addition, Næss (2004: 1192) discusses the well known opposition between accusative and partitive found in Finnish and Estonian (see Campbell 1998, Tamm 2005). In Finnish, highly affected and definite direct objects display accusative case, as in (2.14a), whereas less affected and definite direct objects take partitive case, as in (2.14b):

(2.14) Finnish (Finno-Ugric)

- (a) *hän jo-i maido-n.*  
 s/he drink-PST.3SG milk-ACC  
 "S/he drank (all) the milk"
- (b) *hän jo-i maito-a.*  
 s/he drink-PST.3SG milk-PART  
 "S/he drank (some of the) milk" (Kittilä: 2002: 114)

Through these examples, Næss brings into question the view that typical direct object are low in individuation and highly affected at the same time. In the next section, after discussing the alternative view proposed by Næss (2003, 2007), I will try to find a viable solution to this apparent *aporia*.

Before turning to the discussion of what I call Indexing Approach to DOM, it is worth discussing Aissen's (2003) recent formal analysis based on the distinguishing function of DOM, cast within the framework of *Optimality Theory*.

Drawing upon the functional-typological generalisation that "the higher in prominence a direct object is, the more likely it is to be overtly marked" (Aissen 2003: 437), Aissen identifies two prominence scales and a relational scale, which are no more than slightly modified versions of the animacy and definiteness hierarchies and the grammatical relation hierarchy respectively (see Croft 2003: 130):

(2.15) *Relational scale*: Subject<object

(2.16) *Animacy hierarchy*: human<animate<inanimate

(2.17) *Definiteness hierarchy*: pronoun<proper name<definite<indefinite  
specific<indefinite nonspecific

Aissen bases her analysis on the principle of markedness reversal, according to which “what is marked for direct objects is unmarked for subjects and vice versa” (Aissen 2003: 438). Her model predicts that if a direct object on the hierarchies is (differentially) case marked, then all other direct objects which occur higher on the hierarchies are likewise case marked. On the contrary, agents ranked high on the hierarchies are usually not overtly coded, whereas a subject low in animacy and definiteness is likely to be overtly coded. These functional typological findings are therefore formalised by means of harmonic alignment scales (see Aissen 2003: 440 ff. for a discussion of the OT machinery and references). These hierarchies are meant to capture the relative markedness of possible associations with the relative degree of animacy and definiteness (Aissen 2003: 443). The associations on the left-hand of the hierarchies are less marked than those on the right-hand:

(2.18) **Animacy**

- Su/Hum>Su/Anim>Su/Inan (where >=unmarked)
- Oj/Inan>Oj/Anim/>Oj/Hum

(2.19) **Definiteness**

- Su/Pro>Su/Pn>Su/Def>Su/Spec>Su/Nspec
- Oj/NSpec>Oj/Spec>Oj/Def/>Oj/PN>Oj/Pro

By reversing these hierarchies, Aissen derives the universal constraints that disfavour the above associations of subjects and direct objects with different positions of the animacy and definiteness hierarchies.

(2.20) **Animacy**

- \*Su/Inan>>\*Su/Anim>>\*Su/Hum (where >=unmarked)
- \*Oj/Hum>>\*Oj/Anim>>\*/Oj/Inan

(2.21) **Definiteness**

- \*Su/Nspec>>\*Su/Spec>>\*Su/Def>>\*Su/Pn>>\*Su/Pro
- \*Oj/Pro>>\*Oj/Pn>>\*Oj/Def>>\*Oj/Spec>>\*Oj/Nspec

These harmonic alignment scales represent that languages tend to avoid human and definite direct objects. Similarly, inanimate and indefinite subjects are avoided. Since these combinations of grammatical relations and features do often come up, languages resort to overt morphological marking, such as DOM, to signal the markedness status of the above associations. Aissen then introduces a constraint which penalises the absence of overt case marking for a direct object high in animacy and definiteness (Aissen 2003: 447):

(2.22) \*<sub>C[ase]</sub> : "Star Zero": Penalizes the absence of a value for the feature CASE.

This constraint is then linked via local conjunction (i.e. the combination of two constraints into one, see Smolensky 1995) to the constraints described above. According to Aissen (2003: 448) local conjunction makes it possible to link semantic markedness (represented by the markedness hierarchies) to morphological markedness (expressed by \* constraint):

(2.23) *Local conjunction of Star Zero and animacy*: \*Oj/Hum & \*Case>>\*Oj/Anim & \*Case>>\*Oj/Inan & \*Case

(2.24) *Local conjunction of Star Zero and definiteness*: \*Oj/Pro & \*Case>>\*Oj/PN & \*Case>>\*Oj/Def & \*Case>>\*Oj/Spec & \*Case>>\*Oj/Nspec & \*Case

Aissen therefore argues that these constraints are to be considered as iconicity constraints, since they favour overt marking for semantically marked configurations. Nevertheless, according to the iconicity constraint, all direct objects would be overtly coded. To avoid this, Aissen introduces an economy constraint, which penalises overt morphological marking for all direct objects. This constraint can be inserted at any point of the hierarchies above, thus blocking the overt marking for every kind of direct object on the right (Aissen 2003: 448-449):

(2.25) \*STRUC<sub>C</sub>: penalizes a value for the morphological category CASE.

Aissen then claims that DOM is a typical example of the tension between these two different motivations of iconicity and economy by pushing languages in different directions. Iconicity favours morphological overt marking, while economy disfavours it. Finally, in Aissen's model, the ranking of constraints sorts out the different cut-off points found in DOM in a particular language. For instance, cases where DOM is optional can be dealt with through the insertion of a constraint among other constraints.



Aissen's OT approach has been very influential in recent studies on DOM, especially those from a formalist perspective. Nonetheless, her approach suffers from some serious limitations, both theoretical and an empirical.

First, Aissen's OT account of DOM is no more than a formalisation of well-known functional-typological generalisations adapted to the formal apparatus of OT (see Haspelmath 2008 for a critique from a functionalist perspective and Newmeyer 2005: 222-225 for a generative-based critique). That is, Aissen's model does not seem to add anything new to our understanding of DOM, and indeed she ignores possible counterexamples to her model. In fact, she assumes that DOM is an "absolute linguistic universal" (Aissen 2003: 439) both synchronically and diachronically, since in her analysis changes in DOM systems derive from the re-ranking of constraints. But this does not seem to be the case. Filimonova (2005: 93) has documented some interesting exceptions to Aissen's generalisations.

For instance, we would expect that direct object pronouns would always be overtly coded as opposed to other kinds of NPs. In Nganasan (Samoyedic, Uralic) we find that the accusative affix is limited to definite patients, while personal pronouns have no inflection at all (see section 6.7). In Washo (Hokan), only 3rd personal pronouns distinguish between subject and object forms, while 1st and 2nd person pronouns follow a neutral alignment, i.e. they are not distinguished (Filimonova 2005: 95). Similarly, Siewierska (2004: 150) discusses a number of languages in which indexation is triggered only by third person pronominal direct objects, whilst first and second person pronominal direct objects remain uncoded. Of course, these situations could be the result of diachronic processes of analogy or phonetic changes, as observed by Filimonova (2005: 98). It should be noted, however, that pronouns usually display quite a conservative behaviour cross-linguistically, as they tend to retain morphological distinctions that are lost in other kinds of nominals. Therefore, even when diachronic explanations are available, these anomalous cases are significant counterexamples to the alleged absolute synchronic universality of DOM claimed by Aissen. Moreover, Sardinian offers interesting and documented counter-evidence to the claim that in languages with DOM, direct object nominals on the upper end of the hierarchies in (4.3) and (2.17) will be the more likely to be overtly coded, such as pronouns. Indeed, DOM in Sardinian seems to have started with proper nouns rather than personal pronouns (Putzu 2008).

Secondly, Aissen does not take into account either pragmatic factors, such as topicality, or the indexing function of case marking in her treatment of DOM. Her claim that topicality is especially relevant in cases where case marking is optional (Aissen 2003: 436) is inaccurate and misleading. As we will see in the Chapter 6, information structure plays a key role in the development and the stabilisation of DOM.

Finally, as discussed by Næss (2004: 1200-1201), Aissen claims that her approach in terms of markedness reversal allows her to make predictions not only on properties of direct objects, but also on properties of subjects/agents. That is, Differential Subject Marking (DSM) is taken to be the mirror pattern of DOM, since what is marked for direct objects is unmarked for subjects, and vice versa. Inanimate and indefinite subjects will therefore be overtly coded, since they are not expected to display such properties. The need for disambiguation thus arises. Yet, her predictions on DSM do not appear to be borne out. In some languages, such as Hindi, DSM occurs on highly individuated subjects (see de Hoop and Narasimhan 2008), thus violating Aissen's predictions. In a similar way, in Kambera (Klamer 2008), intransitive subjects may exhibit different case marking based on a variety of factors, such as pragmatic and aspectual properties, the degree of control of the agent over the action, among others. Hence, if DSM (as well as DOM) were based on the need for disambiguating between agents and direct objects of transitive clauses, then the use of DSM on intransitive subjects should not show up in languages, since there is no need for disambiguation, the intransitive subject being the only argument of the clause.

De Swart (2006, 2007) has implemented a Bidirectional OT-analysis of DOM that integrates the Distinguishing and the Indexing approaches. His model is based on the principles of "Minimal Semantic Distinctness" and "Minimal Morphological Distinctness". These principles predict that arguments not minimally distinct in their semantic properties should be morphologically coded, whereas there is no need for morphological encoding when arguments are semantically distinct. In this model, DOM can function as a recoverability or prominence-marking strategy. In the former case, DOM is used to prevent ambiguities that might arise when both arguments display the same properties. De Swart develops an OT-model in which a speaker takes the hearer's perspective into account when s/he chooses whether or not to overtly code a direct object in order to ensure the intended assignment of grammatical functions. In this case, the function of DOM is a distinguishing one. The overt coding of a direct object is not driven by the potential ambiguity between participants, but by its prominent features, e.g. animacy or definiteness. For instance, a speaker can choose to signal that a direct object is definite through the use of overt case marking, as in Hindi and Turkish (De Swart 2007). If s/he chooses not to signal whether or not the direct object is definite, we expect case marking to be absent. De Swart's Bidirectional OT-formalisation of DOM presents various novel aspects vis-à-vis Aissen's model. First, De Swart does not appeal to the notion of markedness reversal to explain DOM. Second, he incorporates into the OT-formalism what I called "indexing" analysis of DOM, although his analysis does not rely upon the notion of transitivity in order to explain DOM patterns. One interesting aspect noted by De Swart is the different role of animacy as opposed to definiteness in DOM systems. While animacy is

the trigger of DOM, definiteness is the result of the presence of DOM. That is, the presence of case marking can influence the definite vs. indefinite interpretation of a direct object, while it has no effect on case marking, since animacy is a lexical feature independent from case marking. It should be noted, however, that in many languages the presence vs. absence of DOM does not contrast between definite vs. indefinite interpretations, as will be shown in Chapter 6. Once again, the absence of information structure parameters in the system designed by De Swart is remarkable.

Some of the problems shown by the Distinguishing Approach, as well as by Aissen's formalisation, have been dealt with by proponents of what I have called the Indexing Approach. In the next section, this approach will be presented and critically discussed, with particular regard to Næss' characterisation of DOM.

### 2.5.2 The Indexing Approach

The distinctive feature of the Indexing Approach is the attempt to relate the use of case marking to the need for highlighting (i.e. indexing) salient semantic and pragmatic properties of the referents of arguments (Siewierska and Bakker 2008: 291, Song 2001: 156). As we have already seen in section 2.2, this view on case marking informs Hopper and Thompson's work on transitivity, as is made clear by the following quotation:

“... It seems to us that the tendency to mark just definite/animate Os reflects *the purer objectness of such Os* (emphasis mine, GI), and simultaneously marks the higher transitivity of the clause as a whole [...]. The facts suggest that there may be a correlation between O case-marking and the cognitive perception of 'prototypical' transitive events” (Hopper and Thompson 1980: 291)

Definite and animate direct objects are deemed “purer” (i.e. prototypical) objects as opposed to indefinite and inanimate direct objects. In this view, therefore, only “genuine” direct objects are case-marked, whereas indefinite and inanimate direct objects are not. Indeed, they are often expressed in formally intransitive constructions or incorporated, as I have discussed in section 2.5.1. Moreover, the marking of animate and definite objects would serve to distinguish not between the syntactic relations of subject and objects, but between prototypical (i.e. animate and definite) and less prototypical members (i.e. indefinite and inanimate) of the class of direct objects.

Recently, Næss (2004, 2007) has proposed a (partially) new explanation for the differential marking of direct objects, based on a redefinition of the concept

of transitivity. The idea put forward by Næss hinges on her “Maximally Distinguished Arguments Hypothesis” (Næss 2007: 28-30), according to which participants of a canonical transitive event have to be maximally distinguished both physically (i.e. they should not be the same entity) and conceptually (i.e. they should be clearly distinct as to the roles they play in the event, cf. Croft’s (1991) and Kemmer’s (1993) definitions of INITIATOR and ENDPOINT):

“A prototypical transitive clause is one where the two participants are maximally semantically distinct in terms of their roles in the event described by the clause” (Næss 2007: 30)

This view implies that the two participants in a transitive event should be clearly distinct from each other both in the “real world” (thus excluding from the transitive prototype, e.g. reflexives, whose participants are not physically distinguished) and linguistically, in that they should be semantically highly individuated and clearly signalled with respect to their structural coding. Again, as put forward by Hopper and Thompson (1980), the agent is defined as the entity that volitionally performs an action upon another participant (i.e. the patient), who in turn is maximally affected by the event. Consequently, Næss (2007: 30) defines the “Maximally Distinguished Arguments Hypothesis” as follows:

“A prototypical transitive clause is one where the two participants are maximally semantically distinct in terms of their roles in the event described by the clause” Næss (2007: 30)

Hence, the key notion for defining transitivity is the maximal semantic distinctness between both the semantic roles and the properties of participants. This distinction is achieved, in Næss’ view, when the two arguments of the clause do not share any defining attribute. In other words, an agent should not have any of the defining properties of the patient, e.g. affectedness, and a patient should not have any of the defining properties of the agent, e.g. control. The primary function of case marking then is “to distinguish between the participants of a fully transitive clause, that is, a clause where there is maximal semantic distinction of arguments” (Næss 2007: 153)<sup>8</sup>. An interesting point raised by Næss concerns the status of transitive clauses with regard to the concepts of markedness and prototypicality. Indeed, the transitive clauses considered as prototypical by the author are those which usually exhibit (differentially) coded direct objects. In order to account for this situation, the author introduces a further distinction

<sup>8</sup>In this model, both the discriminatory and the indexing functions of case-marking are performed at the semantic level: the former is used to distinguish the agent from the patient, whereas the latter is used to index salient properties of the two participants, e.g. volitionality or affectedness.

between “transitive clauses”, which correspond to the prototype presented above, and “two-participant clauses”. Prototypical transitive clauses are “a more marked type of two-participant construction”, since they iconically signal the conceptual and semantic distinctness of participants. They thus do require a high degree of attention on the part of discourse participants. In other words, “two participant clauses” are unmarked or less marked than “prototypical transitive clauses”, as they highlight only one participant, namely the human agent. Put another way, clauses with highly individuated direct objects are more marked because they have two prominent participants. Conversely, less individuated direct objects are encoded in unmarked clauses because they are less relevant from a semantic and pragmatic point of view, being more “incorporated” or “close” (in a pre-theoretical sense) to the predicate (cf. Lazard 1994). Hence, in Næss’ model, individuated direct objects are overtly coded because they are prominent. This prominence is due, according to Næss (2003, 2007), to the high degree of affectedness of overtly coded direct objects: that is, affectedness is the parameter that prompts DOM cross-linguistically. Therefore, what is encoded by DOM are not animacy or definiteness *per se*, but rather the affectedness of the direct object, which is the prototypical property of direct objects. In turn, affectedness is encoded in terms of animacy and definiteness, which provide a straightforward criterion to measure degrees of affectedness. In other words, animacy and definiteness are epiphenomena of a more abstract property, i.e. affectedness. Næss argues that this is why affectedness tends to co-vary with animacy and definiteness.

The explanation provided by Næss for the “apparent contradiction” between her hypothesis and what is postulated within the Distinguishing Approach stems from “the failure to keep apart two distinct levels of analysis” (Næss 2004: 1207), namely transitive semantics on the one hand and case marking phenomena on the other one. The former has to do with the general properties of direct objects as opposed to agents, irrespective of the formal system these arguments are encoded in, while the latter “concerns the way in which objects which do or do not show these properties are formally encoded within a specific kind of structural system, namely accusative case-marking systems” (Næss 2003: 1207). From the point of view of transitive semantics, the only relevant property regarding direct objects is affectedness, and we have no information about their relative markedness when they display particular semantic or pragmatic properties, which usually trigger DOM.

The solution to the issue of the markedness of overtly coded direct objects is found when we take into consideration how transitive semantics is encoded in nominative-accusative languages. Næss claims that overt marking of individuated direct objects is owing to the fact that nominative-accusative languages treat affectedness as the marked property in opposition to control and instigation (Næss 2004: 1208). That is, while affectedness and individuation are *unmarked* prop-

erties for the direct objects at the level of transitive semantics, they are *formally marked* within a nominative-accusative marking system, since these languages treat the initiator of the event (characterised by control and instigation) as unmarked from a formal point of view (see Delancey 1984 for a similar explanation). To wit:

“In accusative languages, the effect of the action on the patient is treated as marked, being singled out with the accusative case, while the initiation of the action is the unmarked phase” (Næss 2003: 1208)

Of course, Næss rightly claims that affectedness is a crucial property for direct objecthood closely interacting with the degree of individuation of the direct object referent (cf. Anderson 1984, Beavers 2006, Dowty 1991, Fillmore 1968, Levin and Rappaport Hovav 2005, *inter alios*). Her approach, however, suffers from some weaknesses.

First, one of the major problems of Næss’ approach is that she considers affectedness a property of arguments, defining it in a way that is very close to the notion of empathy as defined, e.g., by Kuno (1987)<sup>9</sup>. In her view, affectedness is a perceivable and physical change of state or location resulting from an event (Næss 2004: 1205). In order to make clear the link between affectedness and differences in case marking, the author discusses some examples that show how a lower degree of affectedness may lead to a deviation from the prototypical encoding of transitive events. A representative example is the alternation in the encoding of direct object in Icelandic, where direct objects which exhibit some degree of control, typical of agents, are overtly coded with the dative rather than with the accusative, the latter being reserved only for affected direct objects, as shown by the opposition in examples (2.6a) vs. (2.6b). Næss discusses the well-known opposition between partitive and accusative in the encoding of direct objects in Finnish, in which the partitive is used for less affected direct objects, whereas the accusative is reserved to direct objects completely affected by the action (see examples (2.14a, b) above). In this case, the degree of affectedness can be assessed through the “part-whole” relation: a definite, specific entity is much more affected by the action than an indefinite one. Thus, beer in ‘drink beer’ is conceived of as less affected than in ‘drink the beer’ because of the part-whole relation. On this point, however, Beavers (2010: 22) rightly observes that two different notions are at play in such a relation, since changes affecting “small parts of entities might be greater than those that affect more” (Beavers 2010: 22).

Second, as we will see in the next chapters, the approach adopted by Næss proves highly unsatisfactory when it is used to analyse DOM systems depending

<sup>9</sup>Kuno (1987) defines empathy as ‘the speaker’s identification, which may vary in degree, with a person/thing that participates in the event or state he describes in a sentence’..

on animacy or information structure parameters, e.g. topicality. Næss accounts for such systems by invoking a sort of salience, a notion that she herself considers “rather vague” (Næss 2003: 1202). The author here claims that human and animate entities in general are more affected because the effects of an action on these entities are perceptually more important and central to humans rather than to inanimates. As she puts it:

“An effect on a human participant is more likely to impinge directly on the lives of both the human in question and those surrounding him [sic] than an effect on an inanimate object” (Næss 2004: 1202).

She motivates this assertion by comparing the verbs “to kill” and “to break” (Næss 2004: 1202). The event described by a sentence such as “He killed him” can be regarded as more effectful, or more “dramatic” than the event described in “He broke the vase”. Such a description of affectedness, as already noted, appears to be closer to the notion of “empathy” (Kuno 1987) than to the concept of affectedness as is defined in the literature.

Lastly, this characterisation of DOM raises some more general theoretical issues. Næss’ proposal correctly points out the necessity of providing a solution to the apparent contradiction between the Distinguishing and the Indexing approaches, nonetheless her analysis of DOM in terms of prototypicality of the marked direct objects at the semantic level as opposed to their markedness at the formal level fails to capture the whole range of factors influencing DOM on the one hand, and comes into conflict with its own theoretical basis, namely prototype theory.

If the agent and patient in a prototypical transitive clause have to be maximally distinct with respect to their semantic properties, and do not have to share any defining property of the other, a highly individuated direct object would show the same semantic properties as the agent, i.e. animate and definite. In this case, thus, the maximal semantic distinctness would fail. Also, this characterisation is in direct conflict with Rosch’s (1978) proposal that the more prototypical of a category a member is rated, the more attributes it has in common with other members of the category and the fewer attributes in common with members of the contrasting categories” (Rosch 1978: 38). That is, prototypical categories tend to be as contrastive as possible with adjacent prototypical categories (see Croft 1988: 170). Consequently, if we assume that the presence of contrasting features is a fundamental attribute for identifying prototypical categories, in Næss’ model the two arguments of a transitive clause cannot be said to be maximally distinct, since they have in common both animacy and definiteness.

Næss discusses only briefly the issue of frequency of less individuated direct objects as opposed to more individuated ones, since her analysis is centred on

the difficulty of postulating that the less individuated direct objects are indeed not encoded as direct objects in many languages. According to the frequency criterion, as discussed in section 2.5.1, direct object position is usually exploited to introduce inanimate and indefinite participants. Even though prototypicality does not necessarily have to be equated with frequency, nonetheless a strict correlation does seem to exist between these two concepts, at least at the linguistic level. The forms representing the more frequent category will generally have a wider range of distribution and frequency of use, as well as less structural marking, than those of a less frequent category.

For the sake of completeness, before turning to the discussion of the Information Structure approach, I would like to briefly discuss the OT-model proposed by Malchukov and de Hoop (2008), which combines the indexing and distinguishing functions of DOM. In their view, DOM patterns are due to two different violable constraints, seen as competing motivations, namely *DISTINGUISHABILITY* and *IDENTIFY*. *DISTINGUISHABILITY* requires case marking to disambiguate between the two core arguments of a transitive clause (de Hoop and Malchukov 2008: 584). *IDENTIFY* serves to encode specific semantic/pragmatic information about the nominal argument in question via case marking. Interestingly, de Hoop and Malchukov (2008) relate the two above-mentioned different functions of DOM to two different structural types of DOM (de Hoop and Malchukov 2008: 573: ff.). The first type, called "asymmetrical" DOM, involves the alternation between an overtly coded and a zero coded form, as in the case of Persian or Spanish above. The second type, called "symmetrical", involves the alternation between two overt markers. An example of "symmetrical" DOM is the alternation found in Finnish between accusative and partitive see above, which has been described as an alternation due to unboundedness and negation (Beavers 2006, following Kiparsky 2001) and/or affectedness (de Hoop and Malchukov 2008). However, cases of symmetrical DOM seem to be very scarce cross-linguistically. Indeed, the recurrent cases cited in the literature are Estonian and Finnish. Considering the Finnish (as well as the Estonian) alternation between accusative and partitive as a case of DOM is furthermore problematic, since the partitive does not share the distribution of any other case-marking devices (Moravcsik 1978, Luraghi in preparation). Partitives cross-linguistically seem to signal indeterminacy, polarity or unboundedness, rather than indicating the dependency relation between a noun and its head. Partitive case in Finnish is in complementary distribution with other cases, such as nominative (existential subjects), accusative (direct objects), and genitive (complement of adpositions) (see Huumo 2003, 2009, Asbury 2006). In Iemmolo (2011, in preparation) I argue that the symmetrical alternation is not in fact a case of DOM, since in all the cases in which direct object encoding displays an alternation between two overt markers, the parameters at work are completely different from those that usually govern the presence of DOM, such as affected-



ness, aspect, negation, and quantification. I will therefore deal only with cases of asymmetric DOM, which comprises the vast majority of DOM cases cross-linguistically.

After discussing the Information Structure Approach, I will present a possible solution to the issue of the semantic properties of direct objects.

### 2.5.3 The Information Structure Approach

The idea that the information status of the direct object may be the source of DOM systems has been repeatedly proposed in Romance linguistics. For example, Niculescu (1959: 182) defines object marking in Romance as a “modalité syntactique obligatoire destinée à marquer l’objet direct personnel, individualisé (défini, déterminé) et mis en relief du point de vue stylistique”, where “stylistique” should be understood as “pragmatic”. The idea that topicality could be considered the triggering factor for DOM to appear and extend its function has also been claimed by Pensado (1995), who stated that DOM arose from the topicalisation of direct and indirect objects (Pensado 1995: 203).

More recently, Dalrymple and Nikolaeva (2011) have proposed a new model for DOM cast within the *Lexical-Functional Grammar* framework (henceforth, LFG). In their view, DOM is the grammatical coding of the information structure role of *secondary topic*, and represents the grammatical function of primary object (OBJ, see Bresnan and Kanerva 1989, and Vincent and Börjars 2006). When the direct object is a secondary topic, there is a strong tendency to overtly code its pragmatic role grammatically across languages, by means of either verbal affixes (indexation) or case marking. Dalrymple and Nikolaeva’s (2011) account of DOM involves an indexing function as well. In their view, the explicit marking of some direct objects is a means of signalling the “similarities between subjects and direct objects” (Dalrymple and Nikolaeva 2011:15), in that both subjects and objects share the information status of topic, albeit to differing degrees.

The relevance to DOM of semantic properties such as animacy and definiteness is further explained by the fact that these features are sufficient for DOM or secondary topic marking when the construction starts to grammaticalise. As pointed out in the literature, animate entities are more relevant in human discourse than inanimate ones, and new arguments, once they have been introduced to the discourse, increase in definiteness and therefore in topicality (Givón 1984). Before examining Dalrymple and Nikolaeva’s proposal, it is necessary to present some of the basic ideas of their approach, with particular regard to the concept of secondary topic.

### 2.5.4 Secondary topic and DOM

Lambrecht (1994: 5) defines information structure as the component that serves to create a pragmatically structured proposition. A pragmatically structured proposition reflects the speaker's assumption about an addressee's state of knowledge at the time of an utterance. Within information structure, Lambrecht distinguishes two main categories: presupposition and assertion. Presupposition is defined as the portion (or rather, the set of propositions) that a speaker assumes an addressee already knows, while assertion is the proposition "which the hearer is expected to know as a result of hearing the sentence uttered" (Lambrecht 1994: 52). This distinction underlies the concept of sentence topic adopted both in this work and in Dalrymple and Nikolaeva's approach. Lambrecht (1994: 127) defines topic as the referent that the proposition is about. Topic elements are discourse referents about which a speaker asserts something relevant. Usually, they are given information that is prosodically deaccented, identifiable, activated or accessible, definite<sup>10</sup> (Chafe 1976) and within the scope of pragmatic presupposition. The aboutness relation is also due to the contextual topicality (or relevance/salience, see Lambrecht 1994: 55) of a referent in the discourse, which Lambrecht (1994: 55) defines as "the degree to which a referent can be taken to be a center of current interest with respect to which a proposition is interpreted as constituting relevant information". It is worth recalling that Lambrecht distinguishes between topic, which corresponds to a pragmatic relation, and topic expression, which represents the linguistic or grammatical coding of the topic (Lambrecht 1994: 131).

As already proposed by Givón (1984) for ditransitive clauses, and Lambrecht (1994), the topic relation is not obligatorily unique in a sentence. Indeed, more than one referent can be under discussion at the time of an utterance (Dalrymple and Nikolaeva 2011: 75 ff.). We can therefore recognise a primary and a secondary topic within multiple topic sentences with the primary topic being the aboutness topic discussed above. In contrast, the secondary topic is defined in relation to the primary topic. Nikolaeva (2001: 26) defines it as the "entity such that the utterance is construed to be about the relationship between it and the primary topic" (Nikolaeva 2001: 26). Consider the following examples (Lambrecht 1994: 148):

---

<sup>10</sup>Generally speaking, definite NPs are used as topics when a speaker thinks that a hearer is able to identify a referent within a possible range of referents. In other words, definite NPs are often given, available within a register, and identifiable (Givón 1983, Chafe 1976: 39, Lambrecht 1994: 79 ff.). In contrast, indefinite NPs are quite rarely used as sentence topics, insofar as they are more difficult to identify and are usually new, being introduced for the first time into the discourse. Of course, the correspondence between topicality and definiteness is often an imperfect one. For example, while identifiability (one of the main properties of topicality) is presumably a universal category, its grammatical expression, i.e. definiteness, may be lacking in a particular language (Lyons 1999: 279). I will return to this point in the next chapter.

(2.26) a: *Whatever became of John?*

b: *He married Rosa.*

c. *But he didn't really love her.*

In (2.26b), the subject "he" is the topic and the direct object "Rosa" is part of the focus (see Chapter 4 for a definition of focus). In (3c) both the two unaccented pronouns "he" and "her" are topic expressions, but to different degrees. The sentence in (2.26c) primarily conveys information about John (thus being the primary topic), but adds information about Rosa as well by informing the addressee about the relation (already established in (2.26b)) that holds between the two topic referents, in this case that Rosa is not loved by John (Lambrecht 1994: 148). As stated by Nikolaeva (2001: 12), "multiple topics may be ordered with respect to their pragmatic saliency". In (2.26c), the subject is more salient (i.e. topical) than the direct object. This ordering results in the distinction, previously proposed by Givón, between primary and secondary topics<sup>11</sup>. The primary topic can be identified through 'its pragmatic relation to the respective proposition' (Nikolaeva 2001: 27). In contrast, the secondary topic can be defined not only through its relation to the proposition, but also through its relation to the primary topic. These two pragmatic functions are mapped differently onto syntax. Generally, in the unmarked information structure of the clause, the primary topic is closely associated with the grammatical relation of subject/agent, while the direct object exhibits a split, in that it tends to appear in the clause either as part of the focus or as the secondary topic. Nikolaeva further observes that direct objects are singled out as secondary topics only if their relation with the subject is salient enough to be under discussion at the time of the utterance (Nikolaeva 2001: 40). This would explain the split between secondary topic and focus associated with the grammatical relation of direct object. This pattern is consistent with the situation predicted by Givón (1983), who argues that arguments higher up on the case hierarchy are more likely to be topics, while arguments lower down on this hierarchy tend to correlate with focal positions.

As follows from this characterisation of secondary topics, we would expect that in transitive clauses the two participants bear the roles of primary and secondary topic respectively. The agent function is associated with the primary topic role, whereas the direct object function is linked to the secondary topic role. Dalrymple and Nikolaeva (2011) explain this connection, -which is not so strict- as a result of the grammaticalisation of the two roles of topic into agent and direct object functions respectively, following previous proposals by Givón (1984, 2001) on the diachrony of grammatical relations.

---

<sup>11</sup>Givón's definition of direct objects as secondary topics, however, does not imply any difference between overt vs. zero-coded direct objects.

Dalrymple and Nikolaeva assert that in many languages, secondary topics can be morphologically encoded via either indexation or case marking, or can be linked to a specific grammatical function, or both. They exemplify the connection between the information structure role of secondary topic and DOM systems based on data drawn from a number of languages. In Northern Ostyak (Uralic, see Nikolaeva 1999, 2001), for example, the secondary topic role is strictly linked to the grammatical function of direct object and is expressed via indexation. Chatino (Oto-Manguean) has a DOM system. Let us briefly consider these two cases.

Northern Ostyak has DOI which signals the number of the direct object. Contrary to what has been previously assumed, in Northern Ostyak object indexation is not triggered by definiteness, since a definite direct object may or may not trigger object indexation, as in (2.27a) and (2.27b):

(2.27) Northern Ostyak (Uralic, Ugric)

- (a) *ma tam kalaŋ we:l-səm*  
 I this reindeer kill-PST-1SG.SUBJ  
 "I killed this reindeer"
- (b) *ma tam kalaŋ we:l-se-nil-am*  
 I this reindeer kill-PST-DU.OBJ-1SG.SUBJ  
 "I killed these (two) reindeer" (Nikolaeva 2001: 16)

Nikolaeva accounts for the absence of the object indexation marker in (2.27a) and its presence in (2.27b) in terms of topicality status. The direct object of the (2.27a) is not topical, and therefore it does not trigger object indexation. The direct object in (2.27b), however, is a secondary topic, and its topical status triggers indexation. Nikolaeva further corroborates her hypothesis by showing that even personal pronouns and proper names, which are held as inherently definite, may or may not trigger object indexation. If object indexation were governed mainly by definiteness, then we would expect personal pronouns and proper names to trigger object indexation. In addition, Nikolaeva shows that the absence of object indexation correlates quite consistently with focality. Indeed, object indexation is always absent with *wh*-words, as well as the answers to these questions, which are usually linked to focality. As a matter of fact, the presence of object indexation results in the ungrammaticality of the clause, as shown by examples (2.28a) and (2.28b):

(2.28) Northern Ostyak (Uralic, Ugric)

- (a) *mati kalaŋ we:l-es/\*we:l-s-elli*  
 which reindeer kill-PST-3SG.SUBJ/KILL-PST-OBJ.3SG.SUBJ  
 "Which reindeer did he kill?"

- (b) *tam kalaj we:l-es/\*we:l-s-elli*  
 this reindeer kill-PST.3SG.SUBJ/KILL-PST-OBJ.3SG.SUBJ  
 "He killed this reindeer" (Nikolaeva 2001: 17)

Therefore, in Northern Ostyak, object indexation seems to be triggered by the (secondary) topic status of the direct object. Indeed, those 83% of the direct objects that trigger indexation are activated or informationally old (Nikolaeva 2001: 23).

Dalrymple and Nikolaeva analyse the correlation between DOM and the secondary topic role in Zenzontepec Chatino (Carleton and Waksler 2002). Chatino is an Oto-Manguean language with a VSO basic word order. In this language, DOM does not correlate with the animacy or definiteness/specificity of the direct object, since both animate and definite/specific direct objects may be either overtly coded or uncoded by the preposition *jiʔi*, as in (2.29a):

(2.29) Zenzontepec Chatino (Oto-Manguean, Zapotecan)

- (a) *Juan Ø-yuʔu-nto:-yu (jiʔi) Maria*  
 Juan C-have-eye-3sg DOM Maria  
 "Juan recognised Maria"
- (b) *nka-lo-yu jiʔi na kuchilu-uʔ ntu-siʔyu-yu yane*  
 C-remove-TOP DOM DET knife-SPEC P-cut-SPEC neck  
 "He took (his) knife and began to cut (his) throat"

In (2.29b), the agent is not overtly expressed but its primary topic role is indicated on the verb through the topic marker. The direct object "knife" is overtly coded, inasmuch as it has been mentioned in previous discourse, as indicated by the presence of the definite article (see Carleton and Waksler 2002: 159). However, the direct object "neck" is uncoded because it has not been previously mentioned in discourse. Since the primary topic is the agent, the overtly coded direct object should bear the role of secondary topic. However, there does not seem to be independent evidence for the secondary topic status of these direct objects.

In their approach, both DOI and DOM are "indexing" strategies. They are exploited as a means of highlighting the "similarities" between topical agents and topical direct objects, in that both of them tend to be topical, although to a different degree, and overtly coded. That is, languages have the tendency to overtly code topics, whether they are agents/subjects or direct objects, either via agreement or case marking. As I have already discussed, direct objects are just as likely to be secondary topics as they are to be in focus according to Dalrymple and Nikolaeva's approach. When they bear the role of secondary topic, they will be likely to be overtly coded, as opposed to focal direct objects, which instead do not

get overtly coded. Interestingly, Dalrymple and Nikolaeva (2011: 164-167) assert that there is no correlation between the structural markedness of topical direct objects and their functional markedness, since, based on text counts of Ostyak texts, as well as Maslova's (2003) text counts in Yukaghir, topical direct objects are as frequent as focal direct objects. Topical direct objects, thus, have to be considered as functionally unmarked.

Dalrymple and Nikolaeva's (2011) characterisation of DOM, however, fails to account for the full range of variation found in DOM and DOI systems. First and foremost, Dalrymple and Nikolaeva's approach assumes that DOM and DOI work in the very same way. As I will argue in Chapter 3, based on the analysis of data from genetically and geographically diverse languages, DOM and DOI are different phenomena with different functional motivations, and are in fact differently sensitive to referential and information-structural factors.

Second, as we have seen in section 2.5.1, direct objects are usually less animate and definite than subjects/agents. Indeed, whilst it is true that direct objects can exhibit some degree of topicality, nonetheless direct objects that are part of the focal (and not topical) domain appear to be more frequent than topical direct objects, as well as linguistically less likely to be overtly coded. The claim that "topical objects are common in human discourse and formally marked objects are just as frequent in languages with DOM as formally unmarked objects" (Dalrymple and Nikolaeva 2011: 166) does not seem to be empirically grounded. If topical direct objects were as frequent as non-topical ones, we would expect text counts to show that direct objects display the properties prototypically associated with topical elements, such as definiteness, specificity, or at least identifiability. Yet, as we have seen in section (2.5.1), statistical counts reveal that direct objects are overwhelmingly inanimate or indefinite. These properties, of course, strongly correlate with the non-topical status of referents. Dalrymple and Nikolaeva also ignore the fact that in naturally occurring discourse, clauses are often too short to have a primary topic, a secondary topic, and a focus. For instance, Hopper and Thompson (2001) show that the majority of clauses in a corpus of conversational English have only one participant. This is consistent with Dubois' (1987) findings on Sacapultec. He shows that, in transitive clauses, direct object position is usually exploited to introduce new lexical referents into discourse, which normally rarely persist through more than a few successive clauses. Again, the use of full lexical NPs correlates with a low degree of topicality.

From a syntactic and semantic point of view, uncoded direct objects have been described as being part of the verbal domain, in the sense that they strongly depend on the governing predicate, as we have seen in section (2.4). From an information structure point of view, uncoded direct objects belong to the comment (or focus) domain, as noted by Lambrecht (1994) and Maslova (2003), among others. For instance, Maslova's (2003) counts show that, even though direct objects are often

topical, nevertheless the general tendency for direct objects is to be part of the focal domain. Indeed, 97% of focus elements of transitive clauses are direct objects. The presence of topical direct objects in the focal domain, however, is not at odds with the high frequency of direct objects as part of the focal domain. As noted by Lambrecht (1994), the focus domain may contain topical (i.e. presupposed) elements. As will be shown in Chapter 8, in Northern Italian it is possible to find overtly coded direct objects modified by additive, scalar, and exclusive particles such as *(ne)-anche*, *perfino* "even", and *solo* "only". The presence of DOM in these cases could be explained by the fact that these particles apply their meaning to the (contrastive) topic of the clause. In the next chapter, however, I will turn to a possible solution to the problem of markedness of individuated direct objects and will propose a functional model for DOM.

## Chapter 3

# A model for DOM and DOI

### 3.1 Towards a model for DOM and DOI

The approaches I have discussed so far still fail to resolve the issue of the markedness of individuated direct objects and the connection of direct objecthood with transitivity. In my view, the failure to find a viable solution to these issues stems from the failure to keep transitivity separate from direct objecthood on the one hand and case marking from indexation phenomena on the other one.

Both Hopper and Thompson (1980) and Næss (2003, 2007) put more emphasis on the underlying semantics of transitivity rather than its formal realisation. If, as noted by Croft (2003: 175 ff.; to appear) formal transitive behaviour is defined by i) the occurrence of accusative case marking, ii) indexation of the direct object, and iii) presence of a morpheme that explicitly codes the transitivity of the sentence (if any of these strategies exists in the language), then there is no longer a need to lump together cases like those found in German or Icelandic (see section 2.2), in which the second participant of the event is not coded by the accusative case. Instead, we are dealing with different constructions that identify different grammatical relations, of which only the one accusatively coded can be deemed formally transitive. Contrary to what has been commonly assumed, there is no proof for positing a unique grammatical relation of direct object even within a single language, given that both the formal and the functional evidence seems to point out that we are dealing with specific instantiations of a different grammatical categories (or rules).

An even trickier problem is the apparently irreconcilable contradiction between the Transitivity Hypothesis and the Distinguishing Hypothesis, since properties that are considered marked for direct objects within the Distinguishing Approach, are viewed as unmarked within the transitivity approach. The mismatch



between the properties of direct objects in the two approaches is summarised in Table (3.2):

	Uncoded DOs	Overtly coded DOs
<i>Distinguishing approach</i>	low in individuation, affected	highly individuated, affected
<i>Indexing approach</i>	low in individuation, partially affected	highly individuated, affected

Table 3.2: Features of direct objects

This conflict stems from the implicit assumption in all of the literature dealing with both DOM and transitivity that transitive clauses and direct objects have the very same prototype. Whilst Hopper and Thompson (1980) provide solid cross-linguistic evidence that prototypical transitive clauses require their direct objects to be highly individuated, other studies (cf. the statistic counts provided by Dahl 2007, Dahl and Fraurud 1993, Dubois 2003 *inter alios*) provide as much evidence for direct objects to be preferentially associated with low animacy and definiteness/topicality, as we have seen in section (2.5.1). If we continue to postulate that the prototypical direct object is the same as the one that appears in the prototypical transitive construction, then it becomes impossible to explain why direct objects overwhelmingly align with the exact reverse properties. If we instead accept that there is not a single prototype for direct object, i.e. that the typological prototype for a transitive clause does not coincide with the typological prototype for the direct object, then what remains to be identified is the typological prototype for the direct object itself.

In order to identify the typological prototype for direct object, it is necessary to keep case marking and indexation strategies distinct. Indeed, it seems highly reasonable to assume that the grammatical behaviour of direct objects is sensitive to different functional factors. As noted by many scholars (Croft 1988, 2001; Lehmann 1988, Moravcsik 1974, Siewierska and Bakker 2008, among others), case marking is a relational strategy tendentially associated with adjuncts, since the roles of obliques are not easily predictable from the semantic properties of the verb. On the contrary, as the roles of arguments are highly predictable from the semantics of the verb, case marking is usually absent on core arguments. Siewierska (1997) has documented that this is indeed so. In her sample of 237 languages, nearly half of the languages have no case marking on nouns for core arguments. Moreover, she noticed that case marking of the subject of intransitive clauses is strongly disfavoured. In Croft's (1988) terms, this happens because case marking typically denotes "non-obvious" grammatical relations.

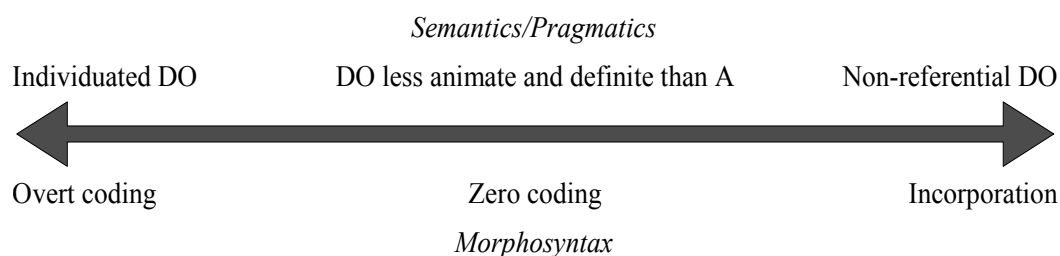


Figure 3.1: Mapping between semantic and morphosyntactic properties of direct objects in DOM

Indexation strategies work the other way round. They are usually exploited to index and keep track of “important or salient arguments” introduced in discourse (Croft 1988: 167; 2001: 235, Givón 1976, 1984; Lehmann 1982; 1988, Siewierska 1997). The reference-tracking function of indexation is confirmed by the fact that, in most of the languages with indexation systems, indexation does not need to co-occur with the NPs they refer to. Hence, indexation strategies are strongly associated with core arguments, and usually shows a preference for nominals high in animacy, definiteness, and topicality. Again, this is confirmed by the data analysed by Siewierska (1997), who shows that indexation is consistently found on core arguments in her sample (and is virtually absent on adjuncts), with a strong preference for the subject of intransitive clauses.

Based on these premises, it follows that two different characterisations have to be provided for DOM and DOI, since they obey different functional motivations, namely the indication of the role of the nominal in the case of DOM and reference-tracking in the case of DOI.

For DOM, a viable solution to the vexing question of the markedness of individuated direct objects could be to claim that direct objects are not prototypically indefinite and inanimate, but rather **less animate and definite** than agents in terms of frequency, as proposed by Comrie (1989: 136). That is, the agent of a transitive clause would simply outrank the direct object in animacy and definiteness, as well as in topicality (see section 2.5.3). This characterisation, summarised in figure (3.1), allows us to explain why highly individuated direct objects get DOM, since they are equal to or outrank the agent in animacy, definiteness or topicality, and therefore need to be clearly signalled. We are also able to account for incorporation phenomena, in which the direct object gets incorporated because it lacks referentiality altogether.

Hence, the morphosyntactic behaviour of direct objects can be described as a *continuum* that runs between two extreme poles, DOM and noun incorporation. The great majority of direct objects, however, are situated in between these two

poles, since the direct object is "the most indefinite position of all major arguments of the verb, in contrast with the subject and dative which are overwhelmingly definite" (Givón 1978: 306).

Givón's and Dahl's frequency counts quoted in section (2.5.1) fully support my proposal that, in the unmarked configuration of semantic properties of agents and direct objects, the agent should not be lower than the direct object in animacy and definiteness as far as its semantic properties are concerned. Moreover, direct objects are more frequent in focal position. Topical direct objects, whether primary or secondary topics, show a clear tendency to receive overt expression, since these direct objects are less typical and therefore more prominent.

As I have already argued in section (2.5.1), differences in animacy and definiteness result from the asymmetry in the roles that participants perform in the transitive event. The initiator of a causal chain (see section 2.4, and Croft to appear: 173) should be volitional and controlling. These properties are more easily ascribed to animate entities than to inanimate ones. From an information structure point of view, these properties make the initiator more prominent as well. The endpoint of the action is instead conceived of as "inert", insofar as it is not volitional and has no control over the action. Accordingly, endpoints of a transmission of force come to be conceptualised as less animate and definite than the agent, as well as less prominent in information structure terms.

Following a proposal by Croft (1988), DOM can thus be assumed to signal the fact that the direct object has "unexpected" (i.e. atypical; Croft 2001: 234) semantic and information-structural properties, in that, contrary to what is usually the case for direct objects, it has a high perceptual and cognitive prominence (Croft 1988: 174; Givón 1985: 206). This is confirmed by the low frequency of animate and definite direct objects in transitive clauses. Nonetheless, as will be shown in the following chapters, it is questionable to assume that the only function of DOM is to distinguish the direct object from the agent at the clause level because both of them are animate and/or definite (as claimed by advocates of the Distinguishing Approach). In point of fact, I will show that DOM in fact correlates with a high degree of topicality of the direct object or with other information-structural differences, such as a high degree of identifiability. Again, the marking of topical or highly identifiable direct objects is brought about by the fact that these properties are not expected to be found in direct objects. Taking into consideration information-structural properties such as topicality makes it possible to explain the degree of variation often found in DOM systems. Furthermore, the restriction of DOM to human and/or definite referents is readily accounted for if one adopts an information structure-based approach, since topicality often involves a conventionalised restriction to human and/or definite entities, as has been convincingly shown by Givón (1976, 1983 and papers therein), Comrie (1989, 2003), and Croft (1988, 2001).

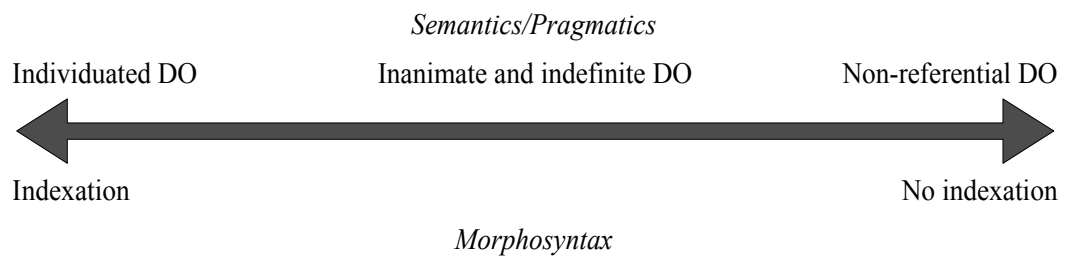


Figure 3.2: Mapping between semantic and morphosyntactic properties of direct objects in DOI

DOI, as opposed to DOM, is correlated with a high degree of topicality of the referents it indexes. In this case, the indexing of animate/definite or topical direct objects is not driven by the unexpectedness of their semantic or information-structural properties. Quite the contrary, it is just because they exhibit these properties that they trigger indexation. DOI can thus be assumed to signal the discourse prominence of the direct object. The discourse prominence is reflected by the fact that, for instance, the NPs that are indexed by DOI are very likely to be left unexpressed because they are highly accessible in discourse (see Lambrecht 1994). Moreover, direct objects that show DOI are often ontologically prominent, since they are human or at least animate (Dahl and Fraurud 1993).

Hence, the morpho-syntactic behaviour of direct objects with regard to DOI patterns in the opposite direction of the behaviour of DOM, as shown in figure (3.1). Put another way, what is a sign of unpredictability for direct objects with respect to case marking (i.e. animacy and/or definiteness and topicality) is predictable for direct objects with respect to indexation (see the notion of relative markedness put forward by Croft 1988).

This characterisation of DOM and DOI as two different constructions allows us to account for the apparent paradoxical behaviour of direct objects. DOM is linked to the unexpected properties of the direct object referents, whilst DOI is systematically associated with the signalling of high salience or prominence of the direct object referents.

So far, we have seen that DOM and DOI are triggered by the very same semantic and information-structural properties, namely topicality and topic-worthiness, the latter being operationalised in terms of animacy and definiteness. I have also argued that, the similarities notwithstanding, the motivations behind DOM and DOI are fundamentally different. DOM signals the unexpected topicality of a referent in direct object position, while in the case of DOI, topicality triggers DOI without being the unexpected property.

The different functions performed by DOM and DOI at the semantico-pragmatic level are mirrored by the different functions they perform within the discourse. As I will show in the following chapters, DOM is overwhelmingly associated with devices that signal discourse discontinuities, such as topic-shift and topic promotion structures. DOI, on the other hand, is consistently associated with highly continuous topics. The evidence for this functional differentiation comes both from synchronic and diachronic facts.

First, as we have widely discussed above, highly topical or topicworthy lexical direct objects are rather infrequent. Their low frequency results in a high degree of unexpectedness. Since their role is not easily predictable, as topical elements are generally introduced in subject position, topical direct objects need to be explicitly highlighted -through DOM-, because of their important role in the discourse. As a matter of fact, they will be further commented upon in the ensuing discourse. By contrast, referents of indexed direct objects are usually highly accessible, as demonstrated by the frequent omission of the direct object in the presence of DOI.

Second, in many languages DOM becomes obligatory, or is restricted to dislocated or topicalised direct objects. As is well known (see discussion in Chapter 4), dislocations and topicalisations are means for putting a constituent in topical position and shifting the topic. By contrast, DOI is a device for encoding the high cognitive accessibility of the indexed referents, thus constituting a reference-tracking strategy (Barlow 1992; Siewierska 2004: ch. 5). For instance, as will be shown in Chapter 6, an independent object pronoun plus DOM rather than just the clitic is used upon a change of topic in typologically and genetically diverse languages (Indo-European, Nilo-Saharan).

Third, there is robust cross-linguistic evidence for the relatedness of object markers to other topic functions. In point of fact, in many languages object markers show a recurrent polysemy with i) frame-setting expressions, like spatio-temporal expressions (as in many Indo-European, Afro-Asiatic, Tucanoan, and Altaic languages); ii) topic and conditional markers (as in some Tibeto-Burman, Nilo-Saharan, and Austronesian languages). In some cases, even if the object marker is no longer polysemous with other topic-related functions, it is still possible to trace it back to a topic marker, as in Galo (see Chapter 6).

Finally, topicality still plays an important role even in languages where it is no longer the primary factor underlying the appearance of DOM and DOI. The importance of topicality can be easily detected when the distribution of DOM and DOI in the so-called "optional contexts" is taken into consideration. For instance, as we will see in Chapters 6 and 8, DOM and DOI in several languages depend primarily upon the semantic features of the direct object, like animacy or definiteness, as, e.g., in Romance and Altaic languages. In these languages, DOM and DOI will be obligatory with certain NP classes, usually those higher on the animacy/definiteness hierarchies, regardless of their topicality. DOM and DOI are

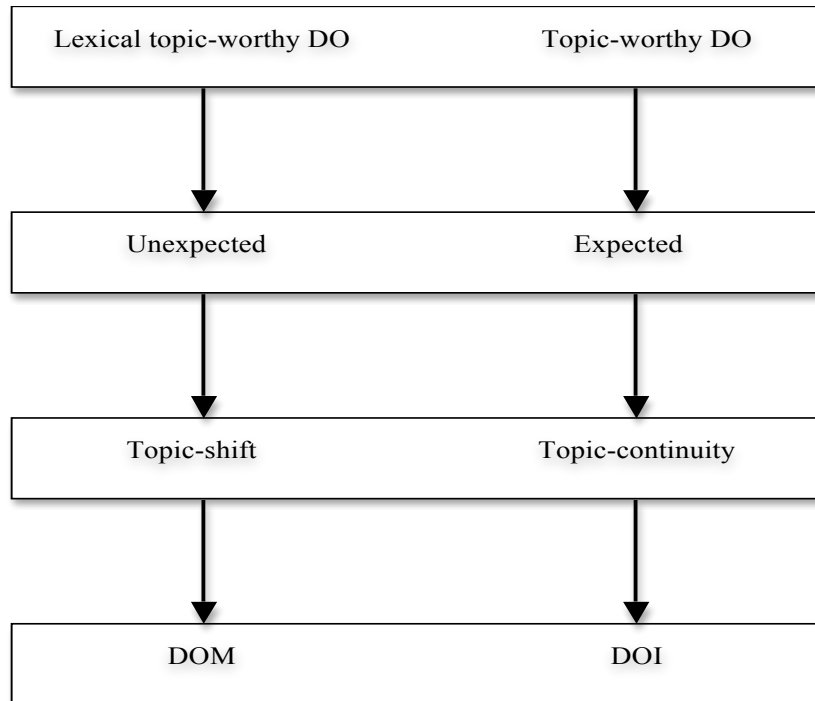


Figure 3.3: Mapping between DOM and DOI, semantics and information structure

instead optional with nominals lower on the hierarchies, such as human common nouns in many Romance languages or indefinite specific NPs in Altaic languages. In the latter cases, topicality often becomes the decisive factor in determining the presence versus the absence of DOM or DOI. That is, the extension of DOM to another NP class applies first to topical direct objects, and is later extended to non-topical ones. Thus, topicality is the factor that paves the way for the extension of the constructions to further NP classes. The differences between DOM and DOI with regard to their triggering factors as well as their discourse functions are summarised in figure (3.1).

## 3.2 Markedness and Distinguishability in DOM systems: a reappraisal

The data that will be presented in the following chapters also brings into question the role of markedness and distinguishability as a factor underlying the distribution of DOM systems cross-linguistically. As we have seen above, in section 2.5.1, many explanations of DOM are based on the need for distinguishing between participants of a transitive clause when the direct object is animate and/or definite. Animacy and definiteness are viewed as marked properties for direct objects, since these direct objects exhibit properties usually associated with agents. As proposed above, I do believe that the development and spreading of DOM systems has to do with the unexpectedness of the direct objects. That is, DOM does not emerge to avoid possible ambiguities that would arise if agent and direct object were both animate and/or definite. This position obscures the role of context and extra-linguistic knowledge in determining language processing and comprehension. Rather, DOM emerges with animate and definite direct objects because these direct objects are more prone to become topics. Thus, DOM operates at the information-structure- and discourse-levels, since it is a means for signalling topic discontinuities.

Based on markedness theory, topical or topic-worthy (i.e. animate and/or definite) direct objects are deemed marked because:

1. they are semantically and conceptually more marked than non-topical or non-topic-worthy direct objects (semantic and cognitive markedness);
2. they exhibit overt coding (formal markedness);
3. they are infrequent in text (textual markedness);
4. they are cross-linguistically overtly coded in contrast to non-topical or non-topicworthy direct objects. Therefore, they should be rarer cross-linguistically (typological markedness).

As I have discussed in Chapter 1, the concept of markedness has been criticised by various scholars (Blevins 2004, Haspelmath 2006, Hume 2004, 2006, in press), who argued that markedness is an epiphenomenon of other factors. Indeed, as pointed out by Blevins (2004), the putative "markedness effects" in grammar can be readily accounted for in terms of diachronic processes and language use. Therefore, markedness as such should be excised from synchronic grammar (Blevins 2004: 313). In Blevins' view, in order to explain the epiphenomenal markedness effects, we need to make direct reference to the factors that drive both language change and language use, without invoking markedness. Likewise,

Hume (2004, 2006) emphasises the role of frequency of use and predictability in explaining markedness patterns (such as the emergence of unmarked vowels in epenthesis).

While I find these views somewhat extreme, I think that the concept of markedness is still very useful if it is not used as an umbrella term for disparate phenomena and factors. In point of fact, the explanation for all the markedness criteria for DOM that I have listed above boils down to frequency of use and expectedness (or predictability, in Hume's terms). As we have seen in section 3.1, topical or topic-worthy direct objects are infrequent in texts. This rarity leads to the unexpectedness of topical/topic-worthy referents in direct object position. When a direct object is highly topical, the need for signalling its topicality and its important status in the ensuing discourse arises. However, this is not due to the fact that a topical direct object is cognitively more marked *per se*. Rather, it is the lower frequency of topical/topic-worthy referents in direct object position that explains why these direct objects are less easy to process. In addition, the bias towards predictability (Hume 2004: 3) envisages that grammatical relations in a clause will be assigned based on the expectations (built up through frequency) of the language user. Since the most frequent situation seems to be one in which agents are human and topical and direct objects are less human and/topical than agents, the reversal of this situation would be likely to be harder to process. Thus, frequency effects themselves can account for what have been referred to as markedness effects so far (Bybee 2001, 2010).

Let us now consider the last markedness criterion, typological markedness. Interestingly, typological markedness brings about a somewhat surprising result. We would expect that DOM, taken as a reflection of a semantically marked configuration that adds cognitive complexity and difficulty to morpho-syntax, would be cross-linguistically rare compared to a system in which all of the direct objects receive the same coding, regardless of their semantic and information-structural properties. However, this does not appear to be the case, as the typological study by Sinnemäki (submitted) shows. Based on a sample of 721 languages, Sinnemäki shows that statistically DOM is universally preferred and diachronically more stable than what he labels COM (consistent object marking). His findings show that, surprisingly, languages seem to prefer DOM, the marked construction, over COM, the unmarked construction. This fact casts further doubt on the interpretation of DOM as a manifestation of markedness, since one would expect the marked structure to be the most unstable and less frequent over time, not the other way round.

In addition, diachronic data does not seem to provide any direct evidence that speakers are driven by the need for distinguishing between agents and direct objects at the clausal level. Instead, the creation of object markers seems to emerge from general pathways of language change, such as extension and reanalysis (Cristofaro 2010, submitted). The extension can take place through processes



of inference, such as metonymy (Hopper and Traugott 1993, Traugott and Dasher 2002) or the reanalysis of the argument structure of particular constructions. For instance, the development of object markers out of serial verb constructions, well attested in Sino-Tibetan languages as well as in a number of West African languages (see Lord 1993 for an overview), is triggered by the bleaching of the event denoted by the first verb in the series that eventually led to the reinterpretation of the verb as a marker of the direct object (see Chapter 7 for further discussion). Similarly, the development of directional and/or locative markers into object markers may be motivated by the conceptual similarity between directional markers and object markers, which crucially share the basic meaning of direction (recall the fact that transitive events require an action to be **transferred** by an entity to another entity). Apart from the basic spatio-temporal meaning that these markers display, another interesting attribute they often seem to share cross-linguistically is an "aboutness" meaning. Indeed, in many languages, object markers are historically related to or polysemous with topic markers, which upholds my idea that DOM is ultimately related to topicality, since the fundamental function of DOM is to overtly signal important participants in discourse (see Chapter 6 for further discussion).

As for the emergence and the development of DOM, we will show that in a number of languages (Indo-European, Tibeto-Burman, Uralic, etc.), DOM seems to have started in a particular context, namely one in which the direct object is dislocated and thus highlighted. DOM, originally restricted to this particular context, then came to be associated with the direct object role, regardless of its position within the clause and its information status. It should be also noticed that in Romance languages the overt coding started off with personal pronouns, which is the only category that retained a case distinction. This pattern will be discussed in Chapter 8.

These diachronic facts, of course, do not mean that the development and the use of DOM are entirely independent from the need for signalling unexpected and infrequent combinations of semantic or information-structural features. However, the data presented here supports a view that sees DOM and DOI as a means to structure discourse in such a way that language users can easily manage the most relevant information and keep track of important participants in discourse. The fact that both DOM and DOI emerge with animate and definite direct objects is due to the fact that these direct objects are more likely to become topics than inanimate and indefinite ones.

An analysis cast solely in terms of markedness and distinguishing function, thus, does not seem to hold either diachronically or synchronically. We have seen that the concept of markedness is not needed to account for either the synchronic distribution or the diachronic development of DOM cross-linguistically. Rather, direct reference is made to frequency and expectedness. In addition, the facts

reviewed here, supported by the data I will analyse in the next chapters, downplay the need to distinguish between arguments within a single clause as the primary motivation behind DOM. Whilst it is indubitable that DOM helps to clarify "who is doing what to whom", this cannot be assumed to be its primary function, since a pure distinguishing function seems to be very rare cross-linguistically, as will be shown in the following chapters. Rather, the exact motivations for DOM and DOI can only be determined when they are considered within the broader context of discourse patterns of transitive constructions. Indeed, the distribution of DOM and DOI is strongly affected by the discourse function of the overtly coded or indexed referent.

As will be shown by the data that will be presented in the following chapters, DOM systems seem to be strongly influenced by economic considerations, insofar as only the infrequent and less predictable combination of semantic and information-structural features brings about the overt coding of direct objects (see Haspelmath 2008, Aissen 2003, and de Hoop and Malchukov 2008 for an OT-formalisation).

Our data, however, makes it clear that DOM systems exhibits a high degree of redundancy as well. As a matter of fact, an interesting aspect that emerges from the data is the frequent co-occurrence of DOM and DOI (as in many Romance and Semitic languages, among others)<sup>1</sup>. Such a co-occurrence, attested in quite a few languages, calls for a further explanation, since it is at odds with the idea that DOM systems are driven by economy. If the overt coding of animate direct objects were chiefly based on the need for discriminating between agents and direct objects in the most economical way, why do we find languages in which both systems are at work? The problem here lies in the considerable emphasis put on the notion of economy, which downplays and obscures the role of redundancy inherent to linguistic systems. Indeed, it seems that linguists often seek to eliminate redundancies in their models (see Croft 2001: 119-124 for a critique of the concept of non-redundancy).

I subscribe to the view that indexation is fundamentally a discourse-driven phenomenon, which serves to relate a syntactic element (be it an NP, a pronoun, etc.) and a discourse referent (see Barlow 1992, 1999). Indexation, as we have discussed above, is a means to index and keep track of salient referent within discourse. This position, however, does not prevent us from recognising the redundant nature of the presence of both DOM and DOI in the very same contexts. How and why does this situation come about?

---

<sup>1</sup>In fact, there are two exceptions to the co-occurrence of DOI and DOM, notably Santali (Austro-Asiatic) and Palauan (Austronesian). In the former case, DOI cannot occur if the direct object is case-marked, while in the latter case DOM and DOI have a different distribution. See Chapter 5 for further discussion.

Before discussing the specific issue of the co-occurrence of DOM and DOI, a preliminary consideration on the role of redundancy is in order. Contrary to the hidden assumption that redundancy in linguistic systems means that a given element is "unnecessary" or "superfluous" in communication, natural languages (as well as artificial ones) do allow a considerable degree of redundancy that chiefly serves to guide the attention of the language user (Chiari 2007). Indeed, as has been argued by Dahl (2004: 12), redundancy is gainfully exploited in linguistic systems to give prominence to "highly unexpected news items" that are worth paying attention to. Since redundancy is closely linked to prominence, Dahl further argues that "it is often difficult to keep redundancy and prominence management apart" (Dahl 2004: 12). This quotation leads us to the core of the problem. In our view, the functional overlapping between DOM and DOI can be easily accommodated if we keep in mind that both of them ultimately serve to highlight the high "prominence" (i.e. topicality) of the direct object. As we have already discussed, nonetheless, there are some differences ultimately related to the general communicative function that case marking fulfils vis-à-vis indexation. Thus, while DOM signals the unexpectedness of the direct object, DOI helps to keep track of the referents. The redundancy of the co-occurrence of DOM and DOI may be explained by the fact that the two constructions have different functions. As is well-known ever since Givón's (1976) seminal paper, indexation constructions arise from the use of anaphoric markers in topic-shift or dislocation constructions (see Siewierska 2004, Gabriel and Rinke 2010 on Spanish, Harris 2002 on Udi, Janse 2008 on Greek, and Tomić 2006 on the Balkan *Sprachbund* languages, among others). As a consequence of increasingly high frequency (Bybee 2010), such dislocated constructions were reanalysed as neutral as to their information status, and the anaphoric markers underwent a process of phonological reduction by which they came to be attached to the verb. Likewise, DOM may arise in topicalisation contexts. Its function, as I have widely argued above, is different, since it is an indicator of topic promotion or shift.

Thus the co-occurrence of the two constructions can easily be accounted for by their different functions in the discourse, namely the indication of topic shift versus topic continuity respectively. This functional differentiation obeys Lambrecht's (1994: 185) "Principle of the Separation of reference and role" (PSRR). According to this principle, speakers tend to avoid structures in which a referent is introduced and commented upon at the same time. In these cases, speakers often resort to dislocation constructions, which allow the establishment of the lexical NP as a topic in dislocated position, resumed by a pronoun or an indexation marker within the clause. Hence, after being lexically referred to, a referent can be encoded as the normal "unmarked" topic expression, i.e. as a pronoun (see also Prince's "simplifying left dislocation", Chapter 4). The differences between DOM and DOI, as well as their co-occurrence, therefore comply with the requirement of

the PSRR, in that DOM serves to introduce the topic which will be later encoded as a normal topic expression through DOI.

### 3.3 DOM, DOI and diachrony

As will be shown in Chapter 6, the high degree of topicality is the crucial parameter for the diachronic development of DOM and DOI (Dalrymple and Nikolaeva 2011, Iemmolo 2010). While for some languages there is historical evidence for the fact that DOM and DOI arose in formal topic-marking constructions, in other languages the available evidence would seem to point towards a more indirect role of topicality. That the role of topicality is indeed crucial even in the latter cases could be readily explained once two possible grammaticalisation paths are taken into account:

1. GENERALISATION: On the one hand, high topicality over time may lead to the generalisation of DOM or DOI to human or animate direct objects, regardless of the actual topicality of their referent. This happens because animacy, along with definiteness and specificity, is a typical feature of topical elements. The role of topicality, nonetheless, can be still recognised in the apparent optionality found, especially in DOM systems, near the cut-off points for the presence of DOM in a particular language.
2. RESTRICTION: In other languages, conversely, DOM and DOI may be restricted to topical direct objects only. In this case, overt coding is found only on topical direct objects.

These two processes will be examined in Chapter 8 on DOM in Romance languages. Before turning to the analysis of the cross-linguistic distribution of DOM and DOI, I will discuss in the next chapter the parameters relevant for the analysis of DOM and DOI, such as animacy, definiteness, and topicality.

# Chapter 4

## Parameters of analysis

### 4.1 Introduction

In this chapter some important methodological issues will be addressed. First, a brief introduction will be made, in which the constructions (i.e. DOM and DOI) that are the object of this investigation will be delimited and the selection of the parameters for a cross-linguistic research will be discussed. Then the individual parameters will be briefly examined. In section 4.3, the "semantic" criteria, such as animacy, definiteness, and specificity, commonly said to trigger DOM and DOI will be presented. It will be shown that an approach to these notions in terms of information structure is more fruitful for a better understanding of the interaction of such properties with DOM. In sections 4.3.4 and ??, the discussion deals with the information structure notions of topic and focus. Finally, section 4.4 details the sample on which this research is based, and summarises the different roles that the various parameters play in the appearance of DOM and DOI.

I will also highlight that, although DOM and DOI in some languages are synchronically constrained by semantic properties such as animacy, where diachronic evidence is available, the role of general information structure parameters is crucial in the development of these constructions.

### 4.2 DOM and DOI: structural variation

As I have mentioned in Chapter 1, animate, definite and/or topical direct objects can be marked either on the head (thus giving rise to DOI) or on the dependent (which is the proper DOM). As far as DOM is concerned, a number of distinctions can be drawn based on the variation found at the structural level. DOM can be expressed by means of affixes (both prefixes and suffixes) as well as adpositions (both pre- and postpositions). Whilst the formal alternations found for DOM do

not correspond to any visible functional distinction, the situation found within the indexation domain correlates with some discourse-functional differences. There has been much discussion about the status of indexation markers vis-à-vis the arguments they index in the literature. Indeed, in many generative studies a neat distinction has been traced between agreement and antecedent-anaphora relation (see Jelinek 1984, Baker 1996, 2008, Bresnan and Mchombo 1986, 1987, among others).

For instance, Bresnan and Mchombo (1987), in their study on Chichewa, differentiate between grammatical and anaphoric agreement. In their view, grammatical indexation marking redundantly signals the features of the argument it indexes and must co-occur with an overt nominal or pronominal argument, which "bears an argument relation to the verb" (Bresnan and Mchombo 1987: 741). In anaphoric agreement, instead, the indexation marker is an incorporated pronominal that stands in an argument relation to the verb. Anaphoric agreement markers, unlike grammatical indexation, do not have to be accompanied by an overt NP.

Siewierska (1999, 2004) has modified Bresnan and Mchombo's (1987) typology of indexation systems by subdividing the grammatical agreement markers into ambiguous and grammatical agreement markers. Ambiguous indexation markers are obligatory but, unlike the grammatical indexation markers, they may occur with or without overt NPs, whether nominal or pronominal. Moreover, ambiguous indexation markers can be distinguished from anaphoric indexation markers in that the latter, but not the former, are in complementary distribution with overt NPs. Siewierska (1999) thoroughly discusses some problems in the identification of the different kinds of indexation markers, such as using complementary distribution as a device with which to distinguish between ambiguous and anaphoric agreement. What is of interest to us here, though, is the claim that, in cases of ambiguous or anaphoric indexation, if an overt NP is present, this does not bear an argument relation to the verb, being rather an "adjunct" to the incorporated pronominal or expressing information structure relations like topic and focus.

That is, the controllers are thus characterised as extra-clausal constituents that do not have argument status, whereas the target is the true argument of the verb, being an incorporated pronominal. This view, quite popular in generative and generative-oriented theories such as LFG (see Bresnan 2001), is however challenged by various facts that I discuss here summarily. Firstly, as many linguists have noted, there are no clear criteria for distinguishing between "true" indexation and antecedent-anaphoric relations on either diachronic or synchronic grounds (Barlow 1988, Corbett 2003, Givón 1976, Lehmann 1988, Moravcsik 1974). Furthermore, as noted by Siewierska (1999) and Croft (2001) the criteria proposed for the identification of the relationship of indexation markers with nominals as arguments as opposed to adjuncts are of limited applicability. For instance, Bresnan and Mchombo (1987) claim that the object indexation marker in Chichewa

has an unambiguous anaphoric function. In their view, the relationship between the indexation marker and the NP is functionally identical to that between dislocated constituents and pronouns in languages like English. They claim that DOI in Chichewa functions solely as an incorporated pronominal argument, and that the NP is co-referential with the indexation marker is a floating topic outside the clause.

The contrast is exemplified by (4.1): in (4.1a) the DOI marker is analysed as a bound pronominal that constitutes the argument of the verb, while the NP is analysed as a clause-external topic.

(4.1) Chichewa (Niger-Congo, Bantu)

- (a) *Mkángo uwu alenje a-na-ú-pez-á*  
 lion this hunter.PL SUBJ-REC.PST-OBJ-find-IND  
*m'nkhalǎngo*  
 in-forest  
 "This lion, the hunters found it in the forest"
- (b) *Paméné á-ná-ú-péza a-na-chéz-á ndi íwo*  
 when SUBJ-PST-OBJ-find SUBJ-REC.PST-chat-IND with it  
 "When they found it, they chatted with it (something other than the lion)" (Bresnan and Mchombo 1987: 770)

In (4.1b), instead, the very same morpheme can be said to work as an indexation marker, since no overt NP is present. In fact, there is no reason to consider the affix *-ú-* in (4.1a) as a bound pronominal and as an agreement marker in (4.1b).

The issue of the argument status of nominals as opposed to that of the indexation markers essentially derives from the assumption that there must be only one syntactic argument per semantic referent in a clause (cf. the "Uniqueness Condition" in LFG: see Bresnan 2001: 47; 91 ff.). In spite of its widespread popularity, nonetheless, this assumption suffers from serious limitations.

First, as we have just seen, it is not clear why one should analyse the same indexation marker as a bound pronominal in one case and as a true indexation marker in the other one. Second, it does not explain why in many languages DOM, which is used to signal a syntactic relation, is found on dislocated elements, which are deemed clause-external. For instance, as will be examined in Chapter 8, DOM in many Romance languages started with (left)-dislocated personal pronouns resumed by a clitic within the clause (i.e. so-called "clitic doubling")<sup>1</sup>. The

<sup>1</sup>Many linguists assume that clitic doubling and indexation must be viewed as completely different phenomena. In my view, however, these two constructions represent different stages along a grammaticalisation cline, with clitic doubling being less grammaticalised than indexation markers (Givón 1979, Creissels 2005, Belloro 2007)

presence of the clitic in Northern Italian shows an interesting range of variation. Indeed, although clitic pronouns co-referential with left-dislocated direct objects have become obligatory elements, in some cases they can be omitted, as shown by (4.2a) vs. (4.2b):

(4.2) Italian (Indo-European, Romance)

(a) A    *me*        *mi*        *preoccup-a*        *il suo*  
 DOM 1SG.OBJ CLIT.1SG worry-3SG.PRS.IND the his  
*comportamento*  
 behaviour  
 "His behaviour is worrying me"

(b) A    *me*        *preoccup-a*        *il suo comportamento*  
 DOM 1SG.OBJ worry-3SG.PRS.IND the his behaviour  
 "His behaviour is worrying me"

If one assumes that the characterisation given above is correct, one would have to analyse the direct object in (4.2a) as clause-external and the one in (4.2b) as clause-internal. It is by no means clear why this should be so.<sup>2</sup> Indeed, this would lead to the conclusion that the preposition *a* marks a core syntactic relation in one case and an adjunct in the other one. This is an unwarranted conclusion, since it is motivated only by purely theory-internal reasons, such as the Uniqueness Condition of LFG.

If indexation markers are instead viewed as discourse-driven constructions, whose chief function is to index and keep track of salient referents in discourse, as I have discussed in Chapter 2, then it makes little sense to claim that there is a syntactic difference in the status of the co-referential NPs between (4.1a) and (4.1b) or between (4.2a) and (4.2b). Furthermore, for my purposes, the presence versus absence of a co-referential NP does not prevent us from giving a unified account of the function of DOI, which is that of highlighting the direct objects because of their high degree of topicality. Indeed, it seems that the appearance DOI is not affected by the structural distinction between anaphoric versus ambiguous indexation. My analysis is empirically corroborated by the fact that the alleged clause-external direct objects indexed by a clitic or an affix may be case-marked (and, indeed, they may be the construction in which case marking develops, as in Italian) and not intonationally marked by a pause, thus becoming integrated into the clause.

<sup>2</sup>For example, as noted by Russi (2008: 110), left dislocations in Italian do not have to exhibit a different intonation pattern, i.e. there is no pause between the dislocated constituent and the rest of the sentence. Even prosodic evidence thus fails to recognise an extra-clausal status for the dislocated element.



This analysis has the advantage of explaining why, for instance, DOM may start or be found in dislocated contexts. It also provides a unified account of the occurrence of DOI, whether or not it is accompanied by an overt NP (see Croft 2001: 226-232 for similar observations).

## 4.3 The parameters

### 4.3.1 Animacy

The notion of animacy strongly permeates many aspects of the grammar of human languages. As is well known, many grammatical constructions are sensitive to animacy. For instance, number marking in many languages of the world is primarily found on animate referents (Corbett 2000). Gender assignment seems to be influenced by animacy as well, as observed by Dahl (2000). As noted by Dahl and Fraurud (1996), however, animacy has received little systematic attention in the literature, a notable exception being the chapter devoted to animacy in Comrie (1989, ch. 9). Animacy can basically be described as an extra-linguistic inherent property of entities, whose manifestation in the grammar of human languages can be represented through the hierarchy given in (4.3); (Comrie 1989, Croft 2003, Dahl and Fraurud 1996):

(4.3) ANIMACY HIERARCHY: human > animal > inanimate

As has been pointed out by many scholars, the conceptualisation underlying the animacy hierarchy is closely related to the anthropocentric and egocentric perspective of the language user (Dahl 2007, Dahl and Fraurud 1996, Yamamoto 1999). One of the explanations provided for the pervasiveness of the animacy hierarchy relies upon the concepts of *empathy*<sup>3</sup> (Langacker 1991) or *point of view* (Dahl and Fraurud 1996). These labels refer to the fact that human beings tend to conceive of the world around us as "organised around animate beings which perceive and act upon their inanimate environment" (Dahl and Fraurud 1996: 60). Hence, events are depicted by human speakers from the point of view of animate beings. It seems clear, thus, that the linguistic manifestations of animacy in grammar do not reflect a simple dichotomy based on the literal sense of the notion of animacy (i.e. a distinction between living vs. non-living entities), but the degree of "salience" that animate entities have in human communication (see Comrie 1989). Indeed, all of the grammatical manifestations of animacy in languages are

<sup>3</sup>The concept of empathy was first proposed by Kuno and Kubaraki (1977), and Kuno (1987) defines it as "the speaker's identification, which may vary in degree, with a person/thing that participates in the event or state he describes in a sentence" (Kuno 1977: 628).

a means to make salient entities clearly stand out because of their importance in discourse or their unexpected occurrence in a particular role.

The quintessential salience of animate entities is reflected in Langacker's decision to replace the term "animacy" with the term "empathy". He justifies his choice by stating that in fact the animacy hierarchy "reflects an egocentric assessment of the various sorts of entities that populate the world. It ranks them according to their potential to attract our **empathy**" (Langacker 1991: 307). He therefore adds a further distinction between speech-act participants (i.e. a person hierarchy), since we primarily tend to empathise with ourselves, as represented in (4.4):

(4.4) EMPATHY HIERARCHY: speaker> hearer> human> animal> physical  
object> abstract entity

Besides the higher likelihood for animate entities to show number marking or gender distinction, animacy plays a key role in pronominalisation, word order and subject/agent vs. object selection. Animate entities show a higher propensity to be pronominalised than to be expressed as full NPs, as demonstrated by Dahl and Fraurud (1996). As for word order, in many languages, animate NPs must precede inanimate NPs. For example, Siewierska (1988: 56) discusses the case of Sesotho, in which the animate NP must precede the inanimate one, as shown by the ungrammaticality of (4.5b) as opposed to (4.5a):

(4.5) Sesotho (Niger-Congo, Bantoid)

- (a) *ke-phehét-sé ngoaná lijó*  
1SG-cook.PST-ASP child food  
"I cooked the child food"
- (b) \**ke-phehét-sé lijó ngoaná*  
1SG-cook.PST-ASP food child  
"I cooked the child food"

The difference in animacy between subjects/agents and direct objects was discussed in Chapter 3. As a matter of fact, animate entities display an overwhelming preference to be selected as subjects or agents, while inanimate entities are more commonly found in direct object position. Finally, animacy has a significant impact on DOM and DOI, which are the phenomena under investigation in this dissertation.

It is very rare, however, to find instances in which animacy alone is responsible for the appearance of a particular grammatical construction. On the contrary, animacy often interacts with other parameters in such a way that makes it impossible,

or at least very difficult, to separate the role of animacy from the role played by other factors, such as definiteness (Comrie 1989: 199). Let us consider DOM: in many languages, animacy is clearly the leading parameter in determining the presence versus absence of DOM. But within the class of human or animate referents that may receive DOM, animacy interacts with other parameters, like definiteness. In (4.6a) from Maithili, the direct object *nokər* is not marked, while it receives DOM in (4.6b). The use of the postposition is not triggered by the animacy of the direct object (or rather, its humanness). On the contrary, the postposition overtly signals that the animate direct object must be interpreted as definite:

## (4.6) Maithili (Indo-European, Indo-Aryan)

- (a) *əhā nokər tək-əit ch-i?*  
 2SG.H servant search-IMPV AUX.PRS-2SG.H  
 "Are you looking for a servant?"
- (b) (*əhā nokər ke tək-əit ch-i?*)  
 2SG.H servant DOM search-IMPV AUX.PRS-2SG.H  
 "Are you looking for the servant?"

Animacy in DOM can also interact with the category of gender. In Russian, DOM is found on personal pronouns and human/animate nouns (Bossong 1998). In the singular, however, only masculine animate direct objects are differentially case marked with the accusative-genitive ending as opposed to the feminine and inanimate ones (Croft 2003: 182):

## (4.7) Russian (Indo-European, Slavic)

- (a) *Marija ubi-la Ivan-a*  
 Marija.NOM kill-PST.SG.F Ivan-ACC  
 "Marija killed Ivan"
- (b) *Marija nash-la kljuch*  
 Marija.NOM find-PST.SG.F key  
 "Marija found the key" (Masha Sotnikova, personal communication)

From the above discussion, it can be seen that the notion of animacy encompasses different intertwining factors rather than animacy in its literal sense. In fact, it has been frequently suggested in the literature that the pervasive role of animacy in human languages is due to the strong tendency for animate entities to be topical. Many studies have pointed out that animate and human referents tend to be more topical both on a sentence- and a discourse level. For instance, the studies collected in Givón (1983) clearly demonstrate that animate referents are

far more often used as topics in discourse than inanimate ones. At the sentence level, it has been repeatedly observed that there is a preference for subjects/agents to be topics. In turn, as we have seen, subjects and agents show a strong tendency to be filled by animate NPs (see Hawkinson and Hyman 1974; Givón 2001: 200, Croft 1991: 151; Siewierska 1988: 65). These facts lend further support to the link between animacy and topicality.

Comrie (1989: 197-199), however, rejects the idea that animacy can be reduced to topicality (he uses the expression "topic-worthiness"), since there seems to be no way of defining topic-worthiness independently of animacy. He cites the case of the distinction between first and second personal pronouns. While there is not any conclusive evidence of their ranking or degree of animacy with respect to each other), first and second personal pronouns seem to differ in their topicality status. That is, a first personal pronoun is usually more topical than a second person, since human discourse tends to be naturally egocentric. However, this view seems to be questionable. First, quantitative data from Cooreman (1987, 1988) on Dyrbal and Chamorro respectively, show that it is impossible to assess a difference between first and second personal pronouns with respect to the Givónian topicality measures of referential distance and anaphoric accessibility (cf. section 2.4). She observes that first and second personal pronouns are "always *given* at any point in the discourse, and hence are always highly topical by the very fact that there is a speaker and a hearer (i.e. an 'I' and a 'you' referent) in every narrative situation" (Cooreman 1988: 95). While this conclusion might appear paradoxical, since it seems to imply that the topicality of first/second person pronouns is so high that it cannot be measured, it highlights exactly the issue at stake here, namely that there does not seem to be any difference in topicality between first and second person pronouns.

Dahl (2000) provides an elegant explanation based on the concept of *egophoricity*. In Dahl's view, supported by statistical counts on adult conversational data from Swedish, English and Spanish, as well as longitudinal data from one Swedish child, the fundamental distinction is between first and second person reference (which he calls egophoric) and third person reference, which he calls allophoric. His findings reveal that egophoric reference, somewhat surprisingly, is more scattered throughout discourse than allophoric reference. The motivation provided for the different behaviour of first/second person reference versus third person reference is based on the different role that these two categories play in discourse. Whereas third person referents must be introduced before talking about them, first and second person referents are the core participants in the discourse, and may be referred to at any point without introduction. Again, first and second person pronouns do not show any difference in their topicality. The purported difference in topicality between first and second person pronouns suggested by Comrie (1989) does seem to be borne out by the data presented by Cooreman's (1987, 1988) as

well as Dahl's (2000) studies. Instead, a difference is found between first/second person as opposed to third person pronouns.

As I have observed above, nonetheless, animacy *per se* is very seldom the sole factor in determining the presence versus absence of DOM or DOI, as well as many other grammatical phenomena. On the contrary, DOM and DOI involve the interaction of independent yet closely related parameters, such as person, definiteness and/or specificity. Since the linguistic reflections of animacy usually involve other parameters as well, such as person and definiteness, it has been suggested to conflate all these parameters in a single hierarchy, the so-called *Extended Animacy Hierarchy* (Silverstein 1976; Croft 2003: 130), which we have already seen in Chapter 4:

- (4.8) EXTENDED ANIMACY HIERARCHY: first/second person pronoun > third person pronoun > proper names > human common noun > non-human animate common noun > inanimate common noun

The effects of the interaction of these different parameters eventually converge in an overall goal, namely that of increasing a referent's salience and therefore making it more likely to show up as a topic in discourse.

### 4.3.2 Definiteness, identifiability and specificity

The notions of definiteness and specificity have been the subject of a strong debate amongst linguists and philosophers of different theoretical persuasions. Many characterisations of definiteness have been provided, and all of them revolve around the semantic-pragmatic concepts of identifiability, familiarity (Karttunen 1976, Heim 1982, von Heusinger 1997, 2002), and inclusiveness (Hawkins 1978 and subsequent works).

In this work, definiteness is taken to be the prototypical grammatical realisation of the semantic-pragmatic notion of identifiability. Following Lambrecht (1994: 75 ff.) and Lyons (1999), I assume that all discourse referents within a discourse can be classified as identifiable or unidentifiable. Identifiability can be defined as "a speaker's assessment of whether a discourse representation of a particular referent is already stored in the hearer's mind or not" (Lambrecht 1994: 76). A referent is thus identifiable when both the speaker and the addressee have a mental representation of that referent, whereas a referent is unidentifiable when only the speaker has a representation of it.

There are different resources according to which the speaker can consider a referent as identifiable by the hearer (Lambrecht 1994: 88 ff.). First of all, a referent can be "inherently" identifiable because there is only one referent that can be designated using that expression. This is the case of proper names, NPs like the

*moon, the Queen, dad*, and generic NPs where an entire class of entities can be referred to with an NP that expresses generalisations about a class as a whole, such as *birds* in *Birds have wings*<sup>4</sup>. Referents may also be made identifiable deictically or anaphorically (Lambrecht 1994: 88-89). In case of deictic referents, the referent is identifiable because it is spatially or temporally near the speech context, and it is thus highly salient within the discourse. For example, inalienable possessed entities, like body parts, are usually signalled as identifiable. In case of anaphoric reference, the referent is identifiable because it has been previously mentioned in the discourse. Usually, once a referent has been introduced into the discourse, it is referred to anaphorically with a pronoun or a definite NP.

Finally, identifiability may be established indirectly through the mention of another entity. When an entity is introduced into the discourse, a frame of identifiable referents is activated. For example, if a war is introduced into the discourse, then the speaker might easily refer to weapons, enemies, etc. as identifiable referents.

The distinctions in identifiability outlined above are presumably universal, since identifiability as a cognitive category is likely to be present as a device for discourse organisation in all of the world's languages. Its grammatical counterpart, i.e. definiteness, has to be kept distinct from the information-structural category of identifiability, as advocated by Lambrecht (1994) and Lyons (1999). That this is a correct assumption is demonstrated by various facts. First, in many languages the formal marking of definiteness may be not fully developed or absent altogether. Second, there is a great amount of variability in the grammatical coding of definiteness across languages; as Lambrecht (1994: 84) observes, the marking of definiteness "should perhaps be seen as reflections of different language-specific cut-off points on the continuum of identifiability". Indeed, the relative unpredictability in the encoding of definiteness could be seen as the result of the grammaticalisation of the category of identifiability (Lyons (1999: 278). This would explain the fact that in some languages some identifiable referents are grammatically treated as indefinite (see the English example above), whereas in other languages referents that are inherently identifiable (and thus definite), like proper names, are overtly marked with the definite article, as in Modern Greek and some Northern Italian varieties.

Definiteness often interacts with the category of specificity in determining the presence of either DOM or DOI. The notion of specificity was developed to account for the behaviour of indefinite NPs, like the NP *a book* in "I am looking for a book". In this sentence, the book may be interpreted as specific or non-specific de-

---

<sup>4</sup>It is worth noting that generics can be overtly coded as definite in some languages. For instance, the Italian translation of the sentence *Birds have wings* is "Gli uccelli hanno le ali", in which *uccelli* as well as *ali* must be obligatorily overtly coded with the definite article.

pending on whether the speaker is looking for a particular book or not (Lambrecht 1994: 80). This ambiguity might be sorted out within the discourse by means of anaphoric reference; if the book is specific, a definite anaphoric expression must be used, as in "I found it/the book". If the book is non-specific, the anaphora must necessarily be an indefinite pronoun or NP, as in "I found one/a book". A useful characterisation of specificity has been put forth by von Heusinger (1997, 2002, to appear). He argues that the core function of specificity is to provide a referential anchoring of the indefinite specific NP to another entity/expression (i.e. the anchor) in discourse (von Heusinger to appear: 43). While the referent of the anchor must be identifiable both by the speaker and the hearer, the referent of the specific indefinite NP must be unknown to the hearer but known to the speaker. On the contrary, referents of non-specific indefinite NPs are not identifiable and anchorable both for the speaker and the hearer (see Lambrecht 1994: 81 for similar observations).

The integration of definiteness and specificity leads to the hierarchy in (4.9):

(4.9) DEFINITENESS-SPECIFICITY HIERARCHY: definite > specific indefinite > non-specific indefinite

This hierarchy (Croft 2003: 132, Givón 1984: 387) is meant to represent the likelihood for a NP higher on the hierarchy to be construed as topical, as well as to be pronominalised. The high probability for a definite NP to be topical lies in the fact that definite NPs are uniquely identifiable by the hearer, who is able to pick up the referent that the speaker has in mind from a range of different referents (see Chafe 1976: 39 ff., Givón 1983: 10). Specific indefinite NPs are less favoured than definite NPs as topics, but still they have a certain degree of identifiability and referentiality that make them possible topics. Indeed, as has been demonstrated by Givón (1989: 179) and Chiriacescu and von Heusinger (2010), specific indefinite NPs have a higher degree of topicality than indefinite NPs<sup>5</sup>. Lastly, non-specific indefinite NPs are strongly disfavoured as topics because they are usually introduced for the first time in the discourse and they are harder to identify. Moreover, as we will see in Chapter 6, in many languages DOM and DOI are confined to definite or specific NPs. Furthermore, even in languages in which animacy takes priority over other parameters, definiteness and specificity are often the parameters that influence DOM besides animacy (see Chapter 6).

### 4.3.3 Activation

A concept closely related to identifiability (and its grammatical counterpart, i.e. definiteness), is activation, i.e. the presence of a referent in the interlocutors'

<sup>5</sup>Other approaches converge to ascribe to specificity properties that are closely related to topicality, such as presuppositionality (Enç 1991)

focus of consciousness or centre of attention. Identifiable referents can be further analysed based on their activation. In this respect, Chafe (1987) distinguishes three degrees of activation in which a referent may be, namely *active*, *semi-active*, and *inactive*.

Active (given) referents are those that are in the focus of interest of the hearer because they have been previously mentioned or are inherently present in discourse. Active referents are generally expressed by unstressed pronouns or anaphoric relations as well as zero anaphora and are not prosodically prominent. However, it is not impossible to find given referents expressed by unaccented lexical NPs (Chafe 1987). Semi-active (accessible) referents are present in the universe of discourse (for example referents evoked in discourse situation) but are out of the focus of interest at the present moment of discourse. Semi-active referents are either referents that have been referred to at some point in discourse and then de-activated or referents that can be inferred from other information that has been already given. Thus, semi-active referents are usually accessible. The sources of such accessibility are various: referents can be textually accessible, situationally accessible, or inferentially accessible (Prince 1981, Lambrecht 1994: 100). Textual accessibility derives from the de-activation from an earlier active state in the discourse, while situational accessibility is due to the presence of the referent in the non-linguistic, external world. Inferential accessibility is usually due to the use of *inferables*, i.e. elements that the interlocutors can infer from information that has been already given or evoked (Prince 1981). For instance, "the driver" in "I got on *a bus* yesterday and *the driver* was drunk" can be inferred from the fact that buses have drivers.

Finally, inactive (new) referents are those that have not been previously mentioned in the discourse, therefore being out of the focus of interest of the hearer. Hence, they tend to be expressed by full NPs and are prosodically prominent, i.e. accented. As Chafe notes, although identifiability and activation are closely inter-related, they should be kept distinct: while the former has to do with the status of referents in the mind of speakers, the latter concerns the status of referents in the mind of the hearer.

Both identifiability and activation are crucial properties for a referent to be construable and interpretable as topical. Indeed, as we will see in the next chapters, DOM and DOI referents are usually identifiable and active or accessible in order to be construed as topical.

#### 4.3.4 Topicality

I have already discussed the notion of sentence topic in section (2.5.4) with respect to Dalrymple and Nikolaeva's approach to DOM. It will suffice here to recall that a sentence topic is the referent that the proposition is about (Lambrecht 1994:



127). Usually it is given information that is prosodically deaccented, identifiable, activated or accessible. Topicality can be characterised at the discourse level as well, as part of discourse-organisation strategies.

The definition of discourse topic overlaps with that of sentence topic: for instance, Ochs, Keenan and Schieffelin (1976: 338) define it as "the proposition (or set or propositions) about which the speaker is providing or requesting information" on a discourse level. Givón's (1983, 1984, 2001) view of topicality overtly equates topicality with discourse topicality<sup>6</sup>. The concepts of topic continuity and referential persistence that I have discussed in section (2.4) are indeed quantitative measures of topicality as a discourse property which is present, albeit in different degrees, in all discourse entities. Referential persistence measures the possibility for an NP to be mentioned again in the ensuing discourse, while topic continuity measures the possibility for an NP to remain or be made the topic of discourse. Discourse and sentence topics are related to each other, in that the choice of a sentence topic is determined by the discourse context and vice versa. As observed by Lambrecht (1994: 127), a referent becomes a topic "if IN A GIVEN DISCOURSE the proposition is construed as being ABOUT this referent" (emphasis in the original). This assertion entails that a sentence topic can be identified as such only when discourse evidence is available. Indeed, all the criteria used to establish sentence topicality go far beyond a single sentence. For instance, the likelihood for an NP to be pronominalised, or the higher tendency for definite and/or animate NPs to be topical can be assessed only on a discourse level.

The apparent dichotomy between discourse and sentence topicality is therefore difficult to determine. However, the property of topicality as a whole is reflected in DOM and DOI systems, though in different ways (see Chapter 3). In fact, whilst DOI systems index the topicality of the DO, and serve to signal topic-continuity, in DOM systems topicality constitutes the unexpected property that triggers case marking. As a matter of fact, in many languages this unexpectedness is mirrored in the function that DOM has at the discourse level. As I will argue in Chapter 6, DOM is consistently associated with a topic-shift, topic promotion function, as is demonstrated by connection between DOM and specific topic-marking constructions or properties, such as dislocations or topicalisations. Examples of such languages are Northern Italian as well as other Romance languages (see Chapter 8). In other languages, on the contrary, DOM is not primarily associated with topicality, but rather with what I call, following Thompson (1990)

---

<sup>6</sup>To wit: "The grammatical organisation of topicality [...] has tended to suggest that "topic" is a clause-level function [...]. This is an unfortunate confusion between the facts of the code of what is being coded. At the level of single event/state, "topic"- "what is talked about" or "what is important" is meaningless. This follows by definition from the two fundamental properties that make individual propositions into coherent discourse, i.e. the fact that human discourse is both multi-propositional and thematically coherent" (Givón 1992).

"topic-worthiness", since overtly coded direct objects consistently display features that are typical of topics across languages, such as animacy, even if they are not really topical. This type of DOM is the result of a diachronic process by which topicality becomes conventionalised in terms of animacy. Another attested situation is when both systems may be at work at the same time, as in languages in which DOM is related to both specific topic-marking constructions and direct objects that are topic-worthy from a discourse level point of view, as in Hindi and Hup (see Chapter 4 for further discussion). It should be noted, however, that in many languages the grammaticalisation process has gone even further, leading to the extension of DOM to direct objects that are not topical, such as *wh*- words. This happens because DOM comes to be associated with the semantic features that are usually associated with topicality, like animacy and or definiteness. Thus, DOM and DOI, though driven by the same information-structural and semantic parameters, perform different functions in discourse.

### 4.3.5 Focality

The notion of focus has been often equated with the concept of "new information" syntactically and prosodically prominent within a sentence. As we have seen in section 2.5.4, Lambrecht (1994) adopts a relational view of the concepts of topic and focus. Based on the notion of pragmatic presupposition and pragmatic assertion (see section 2.5.4), he defines topic and focus as relational concepts that mirror the organisation of a sentence into presupposition and assertion. While the topic of a sentence is presupposed (i.e. already known by the hearer at the time of utterance), focus is the portion of a "pragmatically structured proposition whereby the assertion differs from the presupposition" (Lambrecht 1994: 213). Focus conveys new information, insofar as the relation between the presupposed portion and the assertion is unpredictable for the hearer and therefore non-recoverable (Lambrecht 1994: 207). The newness entailed by the concept of focus lies in the unpredictability of the relation between the focus and the presupposed portion, i.e. the topic. Lambrecht (1994: 221.238) differentiates between three possible kinds of focus structure, namely *predicate focus*, *argument focus*, and *sentence focus*.

The predicate focus is the unmarked type of focus structure (often equated with the topic-comment distinction). It serve to comment on a given topic: in this case, the subject and all other topical elements are presupposed, and the predicate is the assertion that establishes the "aboutness relation between the topic referent and the event denoted by the predicate" (Lambrecht 1994: 226).

(4.10) Q: *What happened to your car?*

A. *My car/It broke DOWN*

The presupposed (hence topical) status of the subject is due to the fact that it can be coded either via a de-accented pronoun or via a de-accented full definite NP. The focus instead, bears the sentence accent that marks the predicate as the focus domain. The pragmatic presupposition of (4.10) can be represented as "speaker's car is a topic for comment  $x$ ", where the comment  $x$  is the assertion fulfilled by the VP "broke down". In predicate focus structures, thus, the focus is constituted by the verb and, in case of transitive verbs, all the constituents inside it. In the argument-focus structure, the focus domain is constituted by a single NP. Its function is to identify the missing argument in a presupposed open proposition, as shown by example (4.11):

- (4.11) Q: *I heard your motorcycle broke down/What broke down?*  
 A: *My CAR broke down*

In (4.11), the presupposition is that "speaker's  $x$  broke down", whereas the assertion is the missing argument "car". Finally, sentence-focus structures (called also "thetic" sentences [cf. Sasse 1987]) lack the pragmatic presupposition altogether, insofar as the assertion extends over the entire sentence, i.e. assertion and focus coincide, as in (4.12):

- (4.12) Q: *What happened?*  
 A: *My CAR broke down*

Except for sentence-focus structures, it has been noted that both argument- and predicate-focus structures may contain presupposed, topical elements (Lambrecht 1994, Nikolaeva 2001). This is not surprising, either from a sentence-based or a discourse-based definition of topic. Indeed, as Lambrecht (1994: 217; 249-251) observes, in predicate-focus structures it is often the verb, or rather the event itself, that constitutes the focus relation, whilst the direct object is topical since it is activated and identifiable. The topical status of the direct object within the focal domain is illustrated by example (4.13):

- (4.13) A: *Has John read Slaughterhouse-Five?*  
 B: *No, John doesn't READ books*

The direct object "books" is topical because of its previous mention of the specific book "Slaughterhouse-Five" in the question. In addition, the NP is deaccented, as opposed to the predicate, which bears prosodic stress. Deaccentuation is indeed a usual means to encode topical constituents (see Lambrecht 1994: 250 ff. for further discussion). Nikolaeva (2001) argues that in argument-focus structures, the direct object may be topical when the focus extends over a constituent other than the direct object, such as adverbials or oblique. In (4.14), for example, the constituent that fills the missing argument slot in the presupposed open proposition "speaker gets the pizza from  $x$ " is the oblique NP "from il Vicino":

- (4.14) A: *Where do you get your favourite pizza?*  
B: *We get it/the best pizza from IL VICINO*

The possibility for topical elements to be part of the focal domain proves very useful for the analysis of DOM as a phenomenon determined by the topicality of the direct object. As a matter of fact, Nikolaeva (2001) and Klumpp (2009, to appear) convincingly demonstrate that DOI in Ostyak and Komi respectively may index direct objects within the focus domain provided that the direct object itself is topical.

### 4.3.6 Dislocation and topicalisation

Dislocations and topicalisations are constructions in which a constituent occurs at the left or right edge of a sentence (Foley 2007: 443). From a formal point of view, dislocations and topicalisations differ only for the presence of a resumptive pronominal or anaphoric element in dislocations, which is instead absent in topicalisations, as shown by (4.15):

- (4.15) *That pasta, I haven't tried*

- (4.16) *That pasta, I haven't tried it*

- (4.17) *I haven't tried it, that pasta*

Various proposals have been made to account for the functions of dislocations and topicalisations. I will discuss here only those that are directly relevant to my analysis of DOM and DOI systems.

Basically, dislocations and topicalisations are used to highlight discourse discontinuities. As a matter of fact, dislocations and topicalisations are employed as topic-shift and topic-promotion devices. Givón (1983, 2001) proposes that the function of dislocations and topicalisations is to re-introduce a referent that has been previously mentioned after a gap of absence. Therefore, referents of dislocated or topicalised NPs are usually accessible and definite or specific. As a matter of fact, indefinite non-specific entities are usually not allowed in dislocated position, because they cannot serve as topics (Givón 2001: 265). Indeed, indefinite NPs used in topical positions receive a specific interpretation forced by their being topical (Leonetti 2004 and references therein).

As noted by Givón (1983, 2001), there are some important differences between left- and right dislocations. Referents of right dislocated NPs are usually already active in the discourse, and are indeed the topic under discussion (Lambrecht 1994: 204). Furthermore, as argued by Givón (2001: 268), their function is to "indicate the end of a thematic paragraph, signalling to the hearer the cataphoric

discontinuity of both the referent and the theme". Thus, right-dislocations are used for referents already active in the discourse. From a syntactic point of view, right dislocated constituents never appear without case marking and are usually closer to the predication (Lambrecht 1994, 2001).

By contrast, left dislocations and topicalisations are topic shift and topic promotion devices. As a matter of fact, they are consistently employed to shift the attention from one referent to another previously mentioned referent not currently active in the discourse. Givón (1983) observed that left dislocations are usually "associated with thematic breaks in discourse structure, i.e. they are typically a paragraph initial device" (Givón 1983: 32). Thus, unlike right dislocations, they signal the beginning of a new passage in discourse by switching the topic. Left-dislocations, as well as topicalisations, serve thus to promote a referent to topic position. The use of dislocations and topicalisations, thus, establishes a referent as the topic for the subsequent discourse. Indeed, referents of left-dislocated NPs are usually referentially persistent in the discourse, while they show low anaphoric accessibility, since they have been absent in the discourse (Givón 2001: 265).

Prince (1997) has proposed a more articulated model for the different types of left dislocations. She identifies three basic types of left dislocations, namely SIMPLIFYING LEFT-DISLOCATION, POSET LEFT-DISLOCATION, AMNESTYING ISLAND-VIOLATION LEFT-DISLOCATION. I will disregard here the last type, as it seems to be quite controversial. The first type, SIMPLIFYING LEFT-DISLOCATION, is used "to simplify the discourse processing of discourse new-entities, by removing them from a syntactic position disfavoured for Discourse-new entities and creating a separate processing unit for them" (Prince 1997: 124). The use of simplifying left dislocations serves to render the referents of dislocated NPs discourse-old so that they may be coded as pronouns (i.e. the usual topic expression) in the subsequent discourse.

The second type, POSET LEFT-DISLOCATION, is employed when the referent of the left-dislocated NP is inferentially linked to another element in the discourse in partially ordered set relations (poset), such as "is-a-member-of, is-a-part-of", etc. (Prince 1997: 127-132). Thus, the left-dislocated NPs "one" and "another" in 4.18 are instances of POSET LEFT-DISLOCATIONS, in that they are part of a set, "the three groups of mice":

- (4.18) She had an idea for a project. She's going to use three groups of mice.  
*One*, she'll feed **them** mouse chow, just the regular stuff they make for mice. *Another*, she'll feed **them** veggies. And THE THIRD, she'll feed [*j*] junk food.

The last sentence involves a topicalised NP. In Prince's view, topicalisations are functionally similar to the POSET LEFT-DISLOCATIONS, insofar as they are

used with partially ordered set relations. For my goals, suffice it so say that both left dislocations and topicalisations are means of promoting a NP to topic position: similarly to left dislocations, this promotion is achieved by putting a referent in sentence-initial position. Interestingly, topicalisations resemble instances of focus preposing. Focus-preposed NPs, however, differ intonationally and pragmatically from topicalised structures, insofar as the preposed NP in focus preposing receives the primary stress and cannot be converted into left-dislocations by adding the resumptive clitic, which is instead possible in cases of topicalisation (Lambrecht 2001: 1053).

For instance, as discussed by Cruschina (2010), following Rizzi (1997), for Italian and Sicilian, topicalised NPs, which represent topical information, are not prosodically prominent and are obligatorily doubled by a clitic. By contrast, focus-preposed NPs bear primary stress and are incompatible with the resumptive clitic, as shown by examples in (4.19):

(4.19) Italian (Indo-European, Romance)

- (a) *Le novelle di Pirandello, Giorgio le ha*  
 the novellas of Pirandello Giorgio 3PL.F.OBJ have.3SG.PRS  
*lette tutte*  
 read.PTCP.PST.PL.F all  
 "Giorgio has read all Pirandello's novellas"
- (b) LE NOVELLE DI PIRANDELLO *ha*  
 the novellas of Pirandello have.3SG.PRS  
*letto (non quelle di Dickens)*  
 read.PTCP.PST.SG.M not those of Dickens  
 "It's the novellas by Pirandello that he read, not those by Dickens"  
 (Cruschina 2010: 52)

As we will see in the next chapters, in languages in which preposed elements can function both as topics of foci, DOM and DOI are consistently associated only with the former.

A further distinction to be taken into account regards the notion of contrast. A contrastive relation involves a relation between an entity to be contrasted and (an)other element(s) on a set of alternatives (Krifka 2007, Molnár 2001). Preposed elements, be they topical or focal, can be associated with contrastive readings (Molnár 2001). Crucially, DOM and DOI are overwhelmingly associated with contrastive topics. As we will see in Chapter 6, even though contrastive topics are often encoded differently from the other topic-marking constructions discussed above with regard to prosody, they possess a similar function, namely that of shifting the topic by marking contrast against the active topic.

## 4.4 The sample

Before turning to the analysis of the data in the next chapters, I should describe the language sample used in this study. As is well known, the problem of language sampling in typological studies has been the subject of an intense debate in recent years, because of the central role played by data in typological research (see Dryer 1989, 1992; Rijkhoff *et al.* 1993; Bakker 2010, Cysouw 2003, Maslova 2003). The number of the world's languages is estimated to be around 6-7000, without counting the languages that have either gone extinct or have become another language because of diachronic changes or language contact. It is therefore impossible to analyse a phenomenon in all the languages of the world, also because many languages have yet to be described.

As Rijkhoff *et al.* (1993) and Bakker (2010) *inter alios*, have pointed out, samples may be variously biased. The most recognised types of bias in a language sample are genetic, areal, typological and bibliographic biases. Hence, some methods have been proposed to build a "representative" sample of the world's languages in typological studies. Depending on the goal of the research, as well as the phenomenon to be investigated, two possible kinds of samples may be built for typological purposes: the probability sample and the variety sample. In the former case, the sample is usually designed to comprise as many genetically and geographically unrelated languages as possible. In the latter case, languages are selected in order to construct a genetically and geographically stratified sample, by including all linguistic families, and their branches and subgroups in proportion to the degree of differentiation they display. In both cases, various problems, such as the unreliability of genetic classification and the stability of some phenomena, often make it impossible not to include genetically and areally related languages. For this reason, various statistical techniques have been developed to address these problems (see Rijkhoff *et al.* 1993, Dryer 1989).

The present research is not based on a typologically balanced sample. In the sample of 175 languages used in this study, I aimed at including all major linguistic families, as well as a number of isolates and creoles. Nonetheless, the sample used in this study should be considered a convenience sample, i.e. there is a bibliographic bias in the selection of the languages. Indeed, no specific method was followed as to the choice of the languages to be included. Rather, the availability and accessibility of detailed grammatical descriptions have played an important role in the selection of the languages examined in this study. Owing to this bias, languages of Eurasia are overrepresented in the present sample. Even though the present sample does not allow for making statistical typological predictions (and statistical predictions were not the aim of this study), it should be remarked that the inclusion of a number of related languages makes it possible to detect subtle differences that may pass unnoticed when dealing with large balanced sam-

ples. Related languages have provided us with precious data to investigate the diachrony of DOM, as will be shown in Chapter 5 and 6. Hopefully, it will become clear that the inclusion of closely related languages has been fruitful for the analysis of DOM from a typological perspective. The sample has been analysed mainly by using descriptive grammars. Where available, monographic studies on DOM have been taken into consideration, especially for those languages where DOM has been thoroughly studied. All the sampled languages, along with their linguistic families and subgroups and the sources that have been used for them, are found in Table (A) in the Appendix.

In the next two chapters, the sample will be analysed based on the distinction between DOM and DOI. The discussion will follow the same structure within each chapter. Each construction will be examined with respect to the main parameter influencing the presence of indexation or case-marking, e.g. animacy, definiteness, or specific topic-marking constructions. The interaction between the various semantic parameters and the overall information-structure function of the constructions I have proposed in Chapter 3 will be constantly analysed at the end of each chapter.



## **Chapter 5**

# **A typology of DOM and DOI: animacy and topic-worthiness**

### **5.1 Introduction**

In this chapter, DOM and DOI systems in which animacy synchronically takes priority over the other parameters will be analysed. Where possible, DOM will be analysed based on language families: otherwise, macro-areas will be taken into consideration. As will become clear, DOM systems governed only by a single parameter are quite rare cross-linguistically. Rather, variation (especially in cases in which DOM or DOI seem to be optional) has to do with the interplay between different parameters. Along the way, I will point out tendencies in DOM systems variation that can be better explained by taking into account the influence of topicality on the apparent optionality found in DOM systems. Indeed, as put forward by Dalrymple and Nikolaeva (2011) and Iemmolo (2010), synchronically DOM and DOI may be related to topicality in an indirect way. As I will argue in the following, DOM and DOI systems based on animacy are the result of the diachronic extension and conventionalisation of the constructions from topical direct objects to direct objects that show features usually associated with topics, like animacy. I will also show that, even in animacy-based systems, topicality often constitutes the secondary factor that paves the way for further extensions of DOM.

### **5.2 Afro-Asiatic**

Many Afro-Asiatic languages exhibit both DOM and DOI. Only in a few languages, however, the predominant parameter regulating the phenomena is animacy. As we will see below, the occurrence of both DOM and DOI is often secondarily influenced by the definiteness of the direct objects, as well as by its

topicality. A straightforward case of DOI governed primarily by the animacy of the direct object is found in Alaaba (Cushitic, East-Highlands; Schneider-Blum 2007), where indexation on the verb is found only with animate direct objects, as in (5.1a, b, and c):

## (5.1) Alaaba (Afro-Asiatic, Cushitic)

- (a) *ʔís(i)*                      *ʔissát(a)*              *hikka*              *b-á*  
 PRON3SG:M:NOM    PRON3PL:ABS    DEM2:ABS    place-TN:M:ABS  
*soh-yó-ss(a)*  
 send-3SG:M:PFV-PC3PL

"He sent them there" (Schneider-Blum 2007: 74)

- (b) *mancú*                      *c'ub-éen(i)*  
 person.SGV:M:ABS    knife-TN:F:LOC/INSTR  
*sh-éema-s(i)*  
 kill-3SG:POL:PFV-PC.3SG:M

"Somebody killed a man with a knife" (Schneider-Blum 2007: 142)

- (c) *ʔís(e)*                      *ʔisú*  
 PRON3SG:F:NOM    PRON3SG:M:ABS  
*wokk'ar-tóo-s(i)*  
 beat-3SG:F/3PL:PFV-PC3SG:M

"She beat him " (Schneider-Blum 2007: 171)

In ditransitive sentences, the indexation marker may refer to either the direct or the indirect object. When referring to the direct object, the clitic expresses its definiteness, as shown by example (5.2a). By contrast, when the clitic indexes the indirect object, it seems to convey a benefactive meaning, as in (5.2b). Schneider-Blum (2007), further notes that agreement with the animate direct object exhibits a certain degree of optionality. In fact, neither object is cross-referenced on the verb in (5.2c). It seems plausible that in this case the presence vs. the absence of the pronominal clitic may be influenced also by information structure factors, whose role cannot be ascertained based on the available data.

## (5.2) Alaaba (Afro-Asiatic, Cushitic)

- (a) *hog-o*                      *mann-aakát(a)*    *min-i*  
 work-TN:F:GEN    person-PL-ABS    house-TN:M:GEN  
*ʔannúh(a)*              *soh-yóom-s(a)*  
 father-SG:M:DAT    send-1SG:PFV-PC3PL

"I sent the (specific) workers for the owner of the house"  
 (Schneider-Blum 2007: 179)

- (b) *hog-o*                    *mann-aakát(a)* *min-i*  
 work-TN:F:GEN person-PL-ABS house-TN:M:GEN  
*ʔanníh(a)*                *soh-yóom-s(a)*  
 father-SG:M:DAT send-1SG:PFV-PC3PL  
 "I sent the workers for the benefit of the owner of the house"  
 (Schneider-Blum 2007: 180)

- (c) *hog-o*                    *mann-aakát(a)* *min-i*  
 work-TN:F:GEN person-PL-ABS house-TN:M:GEN  
*ʔanníh(a)*                *soh-yóom-(i)*  
 father-SG:M:DAT send-1SG:PFV  
 "I sent workers for the owner of the house (I fulfilled my duty, what  
 the owner does with them is up to him)" (Schneider-Blum 2007: 180)

Arabic Lebanese (Koutsoudas 1967), as well as some Arabic dialects of the Galilee and Iraqi and Levantine Arabic (Levin 1987, Eleanor Coghill personal communication) display a system clearly based on animacy. The use of the object marker *l* (often *la-li*), which functions also as a dative marker, is found only with animate nouns introduced by the definite article *l*.

It is worth noting that in many Arabic dialects DOM co-occurs with DOI, i.e. overtly coded direct objects co-occur with a co-indexed overt pronominal in the same clause, as in (5.3a), where the direct object is coded by the preposition *la* and further co-indexed by the clitic *-ha* in (5.3a) and *-u* (5.3b).

(5.3) Arabic Lebanese (Afro-asiatic, Semitic)

- (a) *'Ali* *šaaḥ-ha*                    *la* *l-bint*  
 Ali see.PST.3MS-OBJ.3SG.F DOM the-girl  
 "Ali saw the girl" (Koutsoudas 1967: 512)
- (b) *'Ali* *šaaḥ-u*                    *la* *l-walad*  
 Ali see.PST.3MS-OBJ.3SG.M DOM the-boy  
 "Ali saw the boy" (Koutsoudas 1967: 512)

The prime role of animacy becomes even clearer when the occurrence of DOM with question words is taken into account. The *wh*-word *miin* "who, whom" must be overtly coded whenever there is DOI, as shown by example (5.4a) as opposed to (5.4b):

(5.4) Arabic Lebanese (Afro-asiatic, Semitic)

- (a) *la* *miin* *'Ali* *šaaḥ-u?*  
 DOM who 'Ali see.PST-OBJ.3SG.M  
 "Whom did Ali see?" (Koutsoudas 1967: 513)

- (b) \**miin* 'Ali *šaaḡ-u?*  
 who 'Ali see.PST-OBJ.3SG.M  
 "Whom did 'Ali see?" (Koutsoudas 1967: 513)

Arabic Lebanese, thus, constitutes a clear case where DOM and DOI are entirely dependent upon the animacy of the direct object. Animate question words, which represent typical focus elements, trigger DOI, and consequently DOM, because of the animacy feature. The Arabic Lebanese pattern nicely exemplifies the diachronic extension of DOM and DOI to direct objects with features of topics.

In Biblical Aramaic, as opposed to contemporary Aramaic varieties in which DOM is governed by information structure (see Coghill 2010, to appear and Chapter 6), DOM was primarily restricted to human direct objects (Khan 1984: 470), as shown by (5.5a, b). However, topicality seems to have played a role as well. Khan (1984: 470) reports an example (5.5c) in which an inanimate direct object is overtly coded, because it "plays a prominent role within the text", as demonstrated by the fact that the "furnace" was already mentioned in the text, constituting indeed the central referent of the passage after a gap of absence.

(5.5) Biblical Aramaic (Afro-asiatic, Semitic)

- (a) *lə-hakkīm-ē* *bābel* 'al-*əhōbed*  
 DOM-wise.man-PL.of Babylon PROH-destroy.2SG.M  
 "Do not destroy the wise men of Babylon"
- (b) *ḡdayin* 'Aryōk *han'ēl* *lə-Dāniyēl* *qōdām*  
 then Arioch bring.in.3SG.M.PST.PFV DOM-Daniel before  
*malk-ā*  
 king-DEF.SG  
 "Then Arioch brought in Daniel before the King"
- (c) 'ānē *wə-'amar* *lə-mēzē*  
 answer.ACT.PTCP and-command.3SG.M.PST.PFV to-heat.INF  
*lə-'attūnā* *had-šib'ā* 'al *dī* *hazē*  
 DOM-furnace one-seven over that see.PTCP.PASS.3SG.M  
*lə-mēzəy-ēh*  
 to-heat.INF-3SG.M  
 "He spoke and commanded (them) to heat the furnace seven times more than it was wont to be heated" (Khan 1984: 471)

More data would be needed to ascertain the role that information structure played in the distribution of DOM in Biblical Aramaic. Examples like (5.5c), however, call attention to the role of topicality in determining the usage of DOM with inanimate topical referents.

Maltese represents an interesting case among Semitic languages, because of its long contact with Romance languages, most notably Sicilian and, to a lesser extent, Italian. DOM in Maltese is fundamentally regulated by humanness. Pronouns and proper names trigger the obligatory use of the preposition (*li*), as in (5.6a) and (5.6b). Nouns lower in the animacy hierarchy, such as human common nouns and non-human nouns might be optionally overtly coded as in (5.6c) and (5.6d):

## (5.6) Maltese (Afro-asiatic, Semitic)

- (a) *It-tabib bagħat lil-u*  
 DEF-doctor send.PST.3SG.M DOM-3SG.M  
 "It was him the doctor sent"
- (b) *It-tifel rat lil Marija*  
 DEF-boy see.PST.3SG.M DOM Mary  
 "The boy saw Mary"
- (c) *Tereza rat lit-tifel /it-tifel*  
 Therese see.PST.3SG.F DOM.DEF-boy DEF-boy  
 "Therese saw the boy"
- (d) *Tereza rat lill-kelb /il-kelb*  
 Therese see.PST.3SG.F DOM.DEF-dog DEF-dog  
 "Therese saw the dog" (Borg and Mifsud 2002: 35)

Sicilian influence is quite evident when the marking of indefinite human direct objects is considered. Whereas Semitic languages do not allow the occurrence of DOM with indefinite human direct objects, Maltese does (see 5.7a), in the same way as Sicilian (Iemmolo 2010). Moreover, in case of personification, even definite inanimate direct objects can get overtly coded, as shown by (5.7b):

## (5.7) Maltese (Afro-asiatic, Semitic)

- (a) *It-tabib bagħat lil wieħed ragel*  
 DEF-doctor send.PST.3SG.M DOM one man  
*bir-risposta*  
 with-DEF-answer  
 "The doctor sent a certain man with the message"
- (b) *lx-xemx wiegbet lill-qamar*  
 DEF-sun answer.PST.3SG.F DOM-DEF-moon  
 "The sun answered the moon" (Borg and Mifsud 2002: 35)

### 5.3 Australian

All the Australian languages sampled in this study belong to the Pama-Nyungan genus and are typical examples of split ergative systems. Nominals higher on the animacy hierarchy tend to align accusatively, while nominals lower on the animacy hierarchy tend to align ergatively. As we will see shortly, the cut-off points from an accusative system with DOM to an ergative one vary across languages. DOM in these languages is solely based on the animacy of the direct object. For instance, in Arabana-Wangkangum (Hercus 1994), only pronouns and proper names are always overtly coded with the accusative case, as in (5.8a) and (5.8b). NPs denoting humans can be optionally case-marked as in (5.8c) and (5.8d), whereas non-human common nouns are never overtly coded:

(5.8) Arabana (Australian, Pama-Nyungan)

- (a) *wiya-wiya-la-lhuk uka-nha*  
 laugh-laugh-BEN-HIST he-DOM  
 "They laughed at him" (Hercus 1994: 264)
- (b) *Mathapurda-nha Ngatu-thakali-nha nhanhi-ka*  
 old.man-DOM Ngatu-thakali-DOM see-PST  
 "He saw old man Ngatu-thakali" (Hercus 1994: 286)
- (c) *ama-nha waliawa-ma*  
 mother-DOM follow-IMPV  
 "They follow their mother" (Hercus 1994: 67)
- (d) *uka-nha ulyurla pirda-lhuku, kutha-ngaRu-nga, kutha*  
 this-DOM woman kill-HIST water-matter-LOC water  
*iki-ngura partyarna*  
 carry-CONT all  
 "He killed this woman on account of the water, she was carrying the water away, all of it" (Hercus 1994: 264)

In Nhanda (Blevins 2001), DOM is quite similar to the system found in Arabana. Pronouns and proper names are obligatorily coded with the accusative, as in (5.9a, b). Blevins (2001: 49) states that the accusative case marker *-nha* should be considered as a marker of specificity or definiteness. Nonetheless, in all the examples provided in the grammar, the case marker occurs only with animate direct objects. In fact, unlike Arabana, in Nhanda human nouns trigger the compulsory use of accusative, as in (5.9c), where the omission of the accusative would make the sentence ungrammatical. Non-human nominals (i.e. animate) are optionally overtly coded, as in (5.9d), where the accusative occurs twice, since it encodes both the recipient "me" and the direct object "kangaroo":

## (5.9) Nhanda (Australian, Pama-Nyungan)

- (a) *nyini nha-'i ngayi-'nha*  
2PL see-PST 1SG-DOM  
"You saw me" (Blevins 2001: 81)
- (b) *ngayi nha-'i nyini-'nha*  
1SG see-PST 2PL-DOM  
"I saw you" (Blevins 2001: 81)
- (c) *uthu-nggu ala-kanu abarla-nha aja-a*  
dog-ERG that-FOC child-DOM bite-PRS  
"That dog is biting the boy" (Blevins 2001: 50)
- (d) *ingгаа-nha yawarda-(nha)*  
give.IMP.1SG.DO kangaroo-DOM  
"Give me the kangaroo" (Blevins 2001: 50)

In Pitjantjatjara (Bowe 1990), DOM is confined to pronouns and proper names, which are always overtly coded by *-nya*, as in (5.10a, b) while all the other direct objects, whether they have a human referent or not, are left uncoded, as in (5.10c):

## (5.10) Pitjantjatjara Australian, Pama-Nyungan)

- (a) *Tjitji-ngku ngayu-nya nya-ngu*  
child-ERG 1SG-DOM see-PST  
"The child saw me" (Bowe 1990: 11)
- (b) *Tjitji-ngku Billy-nya nya-ngu*  
child-ERG Billy-DOM see-PST  
"The child saw Billy" (Bowe 1990: 10)
- (c) *Billy-lu tjitji nya-ngu*  
Billy-ERG child see-PST  
"Billy saw the child" (Bowe 1990: 10)

Dharumbal (Terrill 2002) exhibits a more unusual system. DOM, signalled by the case suffix *-(n)a* occurs whenever the direct object referent is human (Terrill 2002: 31-32). Similarly, Holmer (1983: 461) states that the accusative is used only with specific and definite animate direct objects. In this case too, however, there seems to be no influence of definiteness or specificity, but the predominant parameter is animacy, as shown by examples (5.11a) and (5.11b):

## (5.11) Dharumbal (Australian, Pama-Nyungan)

- (a) *Ngatha gatarr-(n)a nha-nh*  
 1SG.ERG man-DOM see-NPST  
 "I can see a man" (Terrill 2002: 31)
- (b) *Ngatha ngi-nha bu-nh*  
 1SG.ERG you-DOM hit-NPST  
 "I will hit you" (Terrill 2002: 40)

Similarly, in Wargamay (Dixon 1981) DOM, expressed by the accusative case marker *-ja*, is usually found on pronouns and proper names. There are some occurrences of the case marker with human nouns, as in (5.12a) and even with an animate noun, as in (5.12b). It seems that the accusative marker can be optionally left aside on human nouns, as in (5.12c, d). Note, however, that there is a difference between the two examples probably due to the fact that the predicate is in the imperative mood. In (5.12c), the direct object "white woman" is overtly coded, and the agent is not expressed. (5.12d) shows the opposite situation, in that the direct object is not overtly coded (i.e. it is in the absolutive case), but the agent is overtly expressed:

(5.12) Wargamay (Australian, Pama-Nyungan)

- (a) *ma:Indu banay muyma-ja*  
 man-ERG choke.UNM boy-DOM  
 "The child saw me" (Dixon 1981: 32)
- (b) *guri-galangu yubaymay binbiṛal-ja*  
 eaglehawk-ERG stolen.CAUS-UNM parrot-DOM  
 "The eaglehawks stole the parrots" (Dixon 1981: 32)
- (c) *wiḡia-na ju:nḡa*  
 white.woman-DOM kiss-IMP  
 "Kiss the white woman" (Dixon 1981: 32)
- (d) *wiḡian ju:nḡa ṛjinda*  
 white.woman kiss-IMP 2SG.A  
 "Kiss the white woman" (Dixon 1981: 32)

Thus, in imperative clauses, the presence of DOM would seem to be regulated by the economy principle. DOM is omitted when the agent is present, while it is used when the agent is absent. This pattern seems to be a little bit peculiar, as case-marking on direct objects is often suspended in imperative clauses (see Dixon 1994: 90 on Lardil, and Comrie 1975 on Finnish).



Waga-Waga (Duungijawu) is an extinct language. According to Wurm (1976), the use of accusative marking in Waga-Waga was broader than in other Pama-Nyungan languages. In fact, the accusative marker *nha* was compulsory with pronouns, and its use was optional, but well attested, with human nouns and even some animate non-human nouns such as "dog". Interestingly enough, the accusative marker also appeared on some inanimate nouns such as such as *dadu* "tree" and *guyum* "fire, camp", as in (5.13).

(5.13) Waga-Waga (Australian, Pama-Nyungan)

- (a) *bugin<sup>y</sup>-du* (ŋ)*i-na* *iya:-u*  
 man-ERG 2SG-DOM bite-FUT  
 "The dog will bite you"
- (b) *ŋa-d<sup>y</sup>u* *ŋunam-ma* *n<sup>y</sup>a-ŋi* *guyum-ba*  
 1SG-ERG children-DOM see-PST camp-LOC  
 "I saw the children in the camp"
- (c) *ŋa:m-bu* *n<sup>y</sup>a-ŋi* *guyum-ma*  
 1PL-DU-ERG see-PST camp-DOM  
 "We two saw the camp" (Wurm 1976: 107)

Finally, in Margany and Gunya (Breen 1981), DOM is confined to personal pronouns, as in (5.14a) and (5.14b), the other nominals, whether human or not, being always left uncoded:

(5.14) Margany (Australian, Pama-Nyungan)

- (a) *gunda* *ṇagaṇiya* *ḍanaṇa*  
 yesterday see-PRS-1SG 3PL-DOM  
 "I saw them yesterday" (Breen 1981: 306)
- (b) *gaṭa* *naya* *ḍa:ḍi* *inaḍa*  
 not 1SG see-PRS 2SG.DOM  
 "I can't see you" (Breen 1981: 316)

Summing up, DOM in Australian languages is fundamentally based on animacy, being determined by the split ergative system which is pervasive in these languages. Overt accusative marking is always found on pronouns, and it is highly likely to appear on proper and, more generally, human nouns. At a first glance, the pattern found in those languages seems to be a distinguishing one, based on the economy principle (cf. Malchukov 2008), as in Pitjantjatjara, Margany or Wargamay. Nonetheless, it should be noted that an analysis in terms of ambiguity

avoidance fails to capture the occurrence of overt object marking. Consider example (5.9d) from Nhanda, in which both recipient and theme are overtly coded by the accusative. In this case, an approach in terms of distinguishability is untenable, since the use of the same case marker cannot be said either to disambiguate between NPs or to be economically motivated. The same reasoning applies to examples in which the direct object is lower in animacy than the agent. According to the distinguishing view, the NP higher in animacy would be interpreted as the agent of the clause. If DOM did have only a distinguishing function, then the use of overt accusative marking in examples (5.8c, d) as well as in (5.12b, c) would be unnecessary, since the default interpretation of the clause would see the NP higher in animacy as the agent and the one lower in animacy as the DO. Ambiguity avoidance alone, thus, does not suffice to explain the occurrence of DOM in these cases. It seems plausible that the extension of case marking could be due to the need for highlighting salient participants, as is quite common in DOM systems. This would explain why, for instance in Waga-Waga, case marking may appear on some inanimate nouns. However, the available data makes it impossible to solve this issue here.

## 5.4 Austro-Asiatic and Austronesian

The only Austro-Asiatic language which shows object marking based on animacy in my sample is Santali (Munda, Neukom 2000). Santali has DOI, since animate direct objects are indexed via an affix on the verb. An interesting fact about DOI in Santali is that it cannot co-occur with a co-referential noun overtly coded for case. That is, if the direct object is coded for case, it cannot be indexed on the verb and viceversa, as shown by examples (5.15a), (5.15b, c, and d), where the direct objects are not coded for case but indexed on the verb:

(5.15) Santali (Austro-Asiatic, Munda)

- (a) *idi-me-a-e*  
take-2SG.OBJ-FIN-3SG.SUBJ  
"He will take you along with him"
- (b) *ere-ker-ie-a-kin*  
deceive-PST<sub>1</sub>:TR-1PL.OBJ-FIN-3DU.SUBJ  
"They (two) deceived us"
- (c) *bəbəoric'-ko*      *dal-ked-e-a*  
terribly-3PL.SUBJ strike-PSTPST<sub>1</sub>:TR-3SG.OBJ-FIN  
"They beat him terribly"

- (d) *bəhu-i*                      *əgu-ked-e-a*  
 wife-3SG.SUBJ bring-PST1:TR-3SG.OBJ-FIN  
 "He brought a wife" (Neukom 2000: 98-99)

This syntagmatic alternation is very interesting and seems to be rather rare cross-linguistically to my knowledge. This system does not seem to fit in the typology of indexation proposed, e.g., by Siewierska (1998). In fact, what matters in the case of Santali is not the complementary vs. obligatory distribution of the indexation marker with free nominal or pronominal markers, but rather the distribution of case marking with indexation pronouns. The obligatory absence of case marking for co-indexed nominals applies to the subject/agent and indirect object functions as well. The motivation underlying this system could be an economic one. Indeed, there is no need for overtly signalling arguments both via indexation and case marking, since the distinguishing function that seems to be at play here is achieved by using only one of the two strategies. It should be noted, nonetheless, that in many languages DOM does not seem to follow an economic pattern, since it often co-occurs with clitic-doubling or some kind of pronominal cross-reference, as we have seen in 5.2 with regard to many Arabic dialects. In those case, an explanation in terms of a "pure" distinguishing function, as I will discuss at the end of this chapter, does not fit the available data. Another plausible explanation of the complementary distribution of case marking vs. indexation, could be related to information structure. That is, it is possible that independent pronouns are used as emphatic forms to express contrast or intensification (Siewierska 2004: 67), whereas indexation carries no such an emphatic meaning. Albeit not very common, this process has a close parallel in Irish, where when an overt subject/agent is present, the verb is in the analytic form, i.e. it does not encode any information about person and number of the subject/agent. Instead, if available, when there is no overt subject/agent, the synthetic form of the verb must be used, which encodes person/number features of its subject (McCloskey and Hale 1983: 499 ff.).

Among the six Austronesian languages sampled in this study, two of them, Begak-Ida'an and Palauan (both Malayo-Polynesian) show an interesting DOM system based on animacy. Begak-Ida'an (Goudswaard 2005) is a typical Philippine language with a two-voice system, the Actor Voice (AV) and the Undergoer Voice (UV) that indicates the semantic role of the subject. (Goudswaard 2005; see also Himmelman 2005).

DOM depends primarily on the animacy, or rather the humanness of the direct object. The preposition used to overtly code human direct objects is *nong*, which is also employed to encode locational adjuncts, headless possessives and indirect objects. The use of the preposition *nong* with direct objects shows an interesting range of variation (Goudswaard 2005: 138 ff). In principle, it is compulsory when the human direct object is constituted by a proper name, as in (5.16a). Its appear-

ance is instead optional (but considered as more "stylish") when the direct object is a pronoun or a human common noun, as in (5.16b), in which the use of case marking forces a definite interpretation (Goudswaard 2005: 139). Nonetheless, as noted by Goudswaard, the preposition can be omitted in less careful speech even with proper names.

(5.16) Begak (Austronesian, Malayo-Polynesian)

- (a) *Benson Bəgisud nong Bəssing panow di' Lahad Datu*  
 Benson AV-send DOM Bessing go LOC Lahad Datu  
 "Benson drives Bessing to Lahad Datu"
- (b) *Kəlu' Julia məng-tabang nong ulun miskin*  
 desire Julia AV-help DOM person poor  
 "Julia wants to help the poor"
- (c) *[...]malu' tu akay məngərəra' nong rumo*  
 [...]want too EXIST AV-look.after DOM 3SG  
 "There would be someone looking after her"

DOM in Philippine languages is not confined to Begak only. For instance, it is found in Tagalog as well as in Hiligaynon (Spitz 2002). In these languages, DOM roughly works in the same way as in Begak, as it is employed with animate direct objects governed by predicates in Actor Voice (Himmelman 2005: 148).

Palauan represents an interesting case of direct object encoding. Both DOM and DOI are used to overtly code direct objects. When human direct objects are governed by an imperfective predicate, they trigger DOM via the preposition *er* (Josephs 1975), as shown by examples (5.17a,b)<sup>1</sup>:

(5.17) Palauan (Austronesian, Malayo-Polynesian)

- (a) *Ng omeka er a ngalek a sechelik*  
 3SG eat.IMPFV DOM child my.friend  
 "My friend is feeding a child" (Nuger 2009: 138)
- (b) *A sensei a mengelebed er a re-ngalek*  
 teacher R-hit DOM PL-child  
 "The teacher is hitting the children" (Woolford 2000)
- (c) *A sensei a omes er a re-ngalek*  
 teacher TOP see.IMPFV DOM PL-CHILD  
 "The teacher is looking at the children" (Nuger 2007: 5)

<sup>1</sup>Please note that the agent is topicalised in (5.17c): this kind of construction, however, does not seem to have any visible effect on the distribution of DOM.

DOM		DOI	
IMPERFECTIVE		PERFECTIVE	
human	inanimate	human	inanimate
	singular		
animate	specific	animate	singular

Table 5.1: Distribution of DOM and DOI in Palauan

When the governing predicate is in the perfective aspect, human direct objects trigger DOI on the verb (Josephs 1975, Nuger 2007, Woolford 2000), as shown by examples (5.18a, b, and c):

(5.18) Palauan (Austronesian, Malayo-Polynesian)

- (a) *Mchelebed-ii a ngalek*  
hit-PFV.3SG DEF child  
"Hit the child" (Josephs 1975: 395)
- (b) *Ak mils-a a Droteo er a party*  
I see-PFV.3SG DEF Droteo at DEF party  
"I saw Droteo at a party" (Josephs 1975: 324)
- (c) *Ak mils-terir a retede el sensei*  
I see-PFV.3PL DEF three teachers  
"I saw three teachers" (Josephs 1975: 43)

The basic rule that predicts the occurrence of DOM and DOI respectively seems to be quite straightforward. DOM occurs only in conjunction with imperfective aspect, whereas DOI appears only with perfective aspect, as summarised in (5.4)<sup>2</sup>. Furthermore, the rule predicts also that human direct objects will always be coded via indexation or case marking.

Yet, this latter "rule" is an oversimplification of the real pattern found in Palauan. As a matter of fact, many examples are found in which DOM and DOI respectively encode inanimate direct objects, provided that they are *singular* and *specific*, as shown by examples (5.19):

<sup>2</sup>The gray-coloured cells indicate that DOM and DOI are optional

## (5.19) Palauan (Austronesian, Malayo-Polynesian)

- (a) *Ng menga er a mera a sechelik*  
 3SG eat.IMPFV DOM DEF orange DEF my.friend  
 "My friend is eating a (particular) orange/the orange"
- (b) *Ng mo kol-ii a meradel a sechelik*  
 3SG go EAT-PFV.3SG DEF orange DEF my\_friend  
 "My friend is going to eat an orange/the orange"

Interestingly enough, mass nouns, as long as they are singular and specific, may be optionally overtly coded (5.20a), in which the direct object "water" gets case-marked since it is singular. Note that the presence of the case marker triggers a specific interpretation. In the perfective aspect, however, the very same NP "water" seems to trigger obligatory indexation even though it is not specific, as in (5.20b). In this case, the presence of indexation does not force a specific interpretation as does the use of the object marker in (5.20a)

## (5.20) Palauan (Austronesian, Malayo-Polynesian)

- (a) *Ke millim (er) a ralm er a blil a*  
 2SG drank.IMPFV DOM DEF water at DEF house DEF  
*delak el mechas? Ng mekngit*  
 my\_mother LK old.woman 3SG bad  
 "Were you drinking the water at my grandmother's house? It's bad"
- (b) *Ke ngilelm-ii a ralm er se er do-muchel el*  
 2SG drank-PFV-3SG water at that.time of 1PL.INCL-start LK  
*merael*  
 walk  
 "Did you drink (*the*) water when we started walking?" (Nuger 2009: 139)

Nuger (2007, 2009), who first noted that the overlap between DOM and DOI is not complete as previously stated (cf. Josephs 1975, Woolford 2000), claims that indexation with direct objects in Palauan is not a true DOI system, since it is not sensitive to specificity. Conversely, the marking of direct objects is a real DOM system, since it hinges on animacy, number (singularity) and specificity. Nonetheless, this analysis seems to be biased by the assumption that the cut-off points along the hierarchies for the two different constructions ought to be completely identical, at least within a single language. That DOM and DOI, in spite of the similarities, work differently is showed by the fact that they are constrained to a particular aspect. This is the most compelling evidence for treating the two

constructions as similar phenomena that have however a different scope. Indeed, as discussed in Chapter 3, the similarities notwithstanding, the core functions of DOM and DOI seem to show some fundamental differences. Whilst DOM serves basically to signal the unexpectedness of a referent displaying specific properties such as animacy, definiteness, etc, in the role of direct object in discourse, DOI signals the high discourse prominence of the indexed direct objects.

Besides the complementarity of DOM and DOI, another rare fact about Palauan is the plurality split it exhibits with regard to direct objects (Nuger 2009: 146 ff.). In this case, both DOM and DOI pattern together, since neither of them can be used with non-plural non-human direct objects. Even though number marking often exhibits the same distribution as DOM or DOI, still the overlap between these two grammatical subsystems appears to be quite unusual cross-linguistically. I will discuss the interaction between DOM and plurality splits in section 5.9 on Hup, which displays an interesting intersection between DOM and number marking.

## 5.5 Indo-European and Dravidian

DOM is found in almost all the sub.branches of Indo-European. Contrary to what is commonly reported, DOM is even found in one Celtic (Welsh) and one Germanic language. The only Germanic language displaying DOM is Afrikaans. While the phenomenon is completely absent in Dutch, Afrikaans employs the preposition *vir* (Dutch *voor*, cognate of the English preposition "for"), to overtly code animate direct objects as well as indirect objects and beneficiaries (den Besten 2000), probably as a result of contact with the creoles spoken in that area (see section 5.6). Basically, *vir* codes obligatorily human NPs, as in (5.21a, b, c, d): animate NPs may be optionally overtly coded, according to den Besten (2000):

(5.21) Afrikaans (Indo-European, Germanic)

(a) *Hulle het vir Piet geslaan*  
 they have DOM Pete beaten  
 "They have beaten Pete"

(b) *Ek het nie vir Piet gesien*  
 I have not DOM Pete seen  
 "I haven't seen Pete" (den Besten 2000: 950)

(c) *dat ek gister vir die man/vir Jan/vir hom gaan sien het*  
 that I yesterday DOM the man/DOM Jan/DOM him gone  
 seen have

"that I went to see the man/Jan/him yesterday" (Molnárfi 2004: 336)

- (d) *Jy moet nie vir hom slaan niet*  
 you must not DOM him beat not

"You must not beat him" (Van Helden 1993: 825)

In Nashta and Pomak (two endangered Slavic languages spoken in Greece), DOM is limited to human direct objects (Adamou 2009). Human direct objects are overtly coded by the preposition *na* in Nashta, as shown by examples (5.22a, b). Non-human and inanimate direct objects are usually not overtly coded. Nonetheless, Adamou (2009) claims that they can be overtly coded when personified, as shown by examples (5.22c, d), where the direct objects "goat" and "tree" respectively are personified:

(5.22) Nashta (Indo-European, Slavic)

- (a) *'jisk-at da a-'sfalj-at na 'moma-ta*  
 want-3PL MOD 3SG.F.DOM-take.down-3PL DOM daughter-DEF  
 "They want to take down the daughter"

- (b) *me-'klava-te na 'mene da uzulu'daja*  
 1SG.ACC-put-3PL DOM 1SG.ACC SUB become.mad-1SG  
 "You made me mad"

- (c) *a-'tfui na 'kos*  
 3SG.F.ACC-hear.3SG DOM goat-DEF  
 "He hears the goat"

- (d) *i'lentfe-to gu-'liziŋe na 'dab-at*  
 fawn-DEF 3SG.M/N.DOM-lick.IMPFV.3SG DOM tree-DEF  
 "The fawn licked the tree" (Adamou 2009)

Pomak makes use of the case marker *-a*, the Old Slavic case marker for the genitive-accusative, for human direct objects, as shown by (5.23a, b). Unlike Nashta, overt coding of animate or inanimate direct objects seems to be absent altogether:

(5.23) Pomak (Indo-European, Slavic)

- (a) *Huse'in i'stja Meri'em-a*  
 Hussein.NOM want.AOR.3SG Meriem-DOM  
 "Hussein wanted Meriem"

- (b) *nasra'din-a sa nje 'rukale na 'sfadba-na*  
 Nasreddin-DOM AUX.3PL NEG call.PFV.3PL to marriage-DEF.D  
 "They didn't invite Nasreddin to the wedding" (Adamou 2009)



Let us now turn to the Indo-Aryan languages. The vast majority of the languages spoken in the Indian sub-continent show either DOM or DOI or both, as well as a split-ergative system based on tense. Here we will examine the distribution found in Kashmiri, Maithili, Marathi, and Hindi (Indo-Aryan). While Kashmiri and Maithili restrict DOM to human and animate direct objects, Marathi and Hindi allow the use of the case marker on inanimate direct objects provided that they are definite and/or specific. Kashmiri is particularly interesting because object marking is further regulated by person hierarchy. When the direct object is higher in animacy than the agent, it will be overtly coded. Otherwise, the direct object stays uncoded even if it is human, as shown by (5.24a)<sup>3</sup>. No overt coding is found with perfective tenses, since Kashmiri has a split ergative system based on tense:

## (5.24) Kashmiri (Indo-European, Indo-Aryan)

- (a) *su chu t̄amis par̄ina:va:n*  
 he is.2SG 3SG.M teaching  
 "He is teaching him" (Wali and Koul 1997:87)
- (b) *su vuch-i me*  
 he see-3SG 3SG.DAT  
 "He will see him" (Wali and Koul 1997:156)

Maithili DOM (coded by postposition *ke/kē*, used also for the dative; Yadav 1996) is not permitted on inanimate nouns. It is compulsory with pronouns, kin terms, and proper names, as in (5.25 a, b, and c). In addition, overtly coded direct objects trigger compulsory object indexation, as shown by the examples below (see Stump and Yadav 1988, Bickel, Bisang and Yadava 1996 for a good survey of the intricate Maithili indexation system). Remarkably, singular pronouns use the case marker *-ra*, whereas plural pronouns use *-ra* followed by the plural case marker *səb/səbh* plus the case marker *ke/kē* (Yadav 1996: 109)

## (5.25) Maithili (Indo-European, Indo-Aryan)

- (a) *t̄ō h̄əm-ra dekh-l-ē*  
 2SG.NH 1SG-DOM see-PST-2SG.NH.1SG  
 "You saw me"
- (b) *(əhā) moh̄ən kē b̄əja-u*  
 2SG.H Mohan DOM call-IMP-2SG.H  
 "You call Mohan"

<sup>3</sup>Kashmiri overtly indexes pronominal direct objects. However, when the direct object is first- or third person, no indexation is allowed.

- (c) *tō səsʊr kē gor ləg-əl-hunh*  
 2SG.NH father\_in\_law DOM feet attach-PST-2SG.NH.3SG.H  
 "You greeted (lit. touched the feet of) the father in law" (Yadav 1996: 73 ff.)

Human common nouns (and sometimes names of animals) are optionally overtly coded. The use of the marker is a means to overtly signal that they must be interpreted as definite and/or specific, as in (5.26 a, b; Yadav 1996: 79). It is also worth noting that all direct objects modified by a possessive pronoun (which is encoded by the same case marker as the accusative/dative) must be obligatorily overtly coded, regardless of the animacy or the (in)-alienability of the referent, as shown by (5.26c, d; Yadav 1996: 80):

(5.26) Maithili (Indo-European, Indo-Aryan)

- (a) *əhā nokər tək-əit ch-i?*  
 2SG.H servant search-IMPFV AUX.PRS-2SG.H  
 "Are you looking for a servant?"
- (b) (*əhā*) *nokər ke tək-əit ch-i?*  
 2SG.H servant DOM search-IMPFV AUX.PRS-2SG.H  
 "Are you looking for the servant?"
- (c) *hun-ka janh kē ke jēt-təinh?*  
 3SG.H-ACC/DAT thigh DOM who press-FUT-3SG.NH.3SG.H  
 "Who will press/massage his thigh?"
- (d) *həm to-ra kursi kē ghuskəu-l-iəuk*  
 1SG 2SG.NH-ACC/DAT chair DOM push-PST-1SG.2SG.NH  
 "I pushed your chair"

In Marathi, the accusative postposition *-lā* appears only on human direct objects in non-perfective tenses (being Marathi a split-ergative language). Some inanimate direct objects may take the marker when they denote specific referents as in (5.27c):

(5.27) Marathi (Indo-European, Indo-Aryan)

- (a) *polis tsorā-lā mārto*  
 policeman thief-DOM beat:PRS.3SG.M  
 "The policeman beats the thief" (Pandharipande 1997: 287)
- (b) *mi sudhā-lā bhetto*  
 I Sudha-DOM meet:1SG.M  
 "I meet Sudha" (Pandharipande 1997: 288)

- (c) *to gharānn-ā sadwazto*  
 he house-PL-DOM decorate:3SG.M  
 "He decorates the houses" (Pandharipande 1997: 288)

Hindi has been extensively studied with regard to DOM (including but not limited to Allen (1951), McGregor (1972), Masica (1991), Butt (1993), and Mohanan (1994). It has a system similar to that found in Marathi<sup>4</sup>. Human and animate direct objects are overtly coded by the postposition *-ko*<sup>5</sup>, as shown by examples in (5.28a, b, c). Inanimate direct objects remain uncoded unless they are construed as definite and/or specific, as shown by (5.28d), where the uncoded direct object can be interpreted either as definite or indefinite. When the direct object is overtly coded by *-ko*, it must be invariably interpreted as definite, as shown by (5.28e):

(5.28) Hindi (Indo-European, Indo-Aryan)

- (a) *Ram-ne Mohan-ko dek<sup>h</sup>a*  
 Ram-ERG Mohan-M.DOM see.PFV.M.SG  
 "Ram saw Mohan"
- (b) *vəh əpne kutte-ko bəhut pyar*  
 he self.POSS.M.OBL dog.M.SG.OBL-DOM much love  
*kərtā-he*  
 do.IMPV.M.SG-PRS.3SG  
 "He loves his dog very much" (Kachru 2006: 175)
- (c) *Ilaa-ne ek bacce-ko u<sup>h</sup>aayaa*  
 Ila-ERG one child-DOM lift/carry-PFV.3SG  
 "Ila lifted a child" (Mohanan 1994: 79)
- (d) *Ilaa-ne haar u<sup>h</sup>aayaa*  
 Ila-ERG necklace lift/carry-PFV.3SG  
 "Ila lifted a/the necklace" (Mohanan 1994: 80)

<sup>4</sup>It should be noted that, in perfective tenses, Hindi ergative system displays a "tripartite" system, in which the intransitive subject is uncoded for case, while the agent is coded by the ergative case and the direct object shows accusative coding provided that the semantic conditions for DOM to appear are fulfilled. Another complex issue partially related to DOM is the distribution of indexation. Indexation in Hindi is in complementary distribution with overt case marking. In imperfective tenses, there is indexation only with the subject/agent, which does not bear any case-marking. In perfective tenses, the verb never indexes the agent, which is overtly coded by the ergative case. The verb instead indexes the direct object only if it does not bear any case-marking. If the object shows DOM, there is neutral indexation (i.e. third person masculine singular) (Mohanan 1994).

<sup>5</sup>*-ko* encodes also experiencer subjects and subjects of passive verbs: these usages, however, will not be dealt with here.

- (e) *Ilaa-ne haar-ko u<sup>h</sup>aayaa*  
 Ila-ERG necklace-DOM lift/carry-PFV.3SG  
 "Ila lifted \*a/the necklace" (Mohanana 1994: 80)

The prime role of animacy in determining the appearance of DOM in Hindi is corroborated by the following examples. Whilst human direct objects must be obligatorily overtly coded, DOM with animate direct objects involves a contrast between specificity and non-specificity. Indeed, direct objects construed as non-specific usually show up as uncoded, as in (5.29a) as opposed to (5.29b), where the "cow" is instead specific:

(5.29) Hindi (Indo-European, Indo-Aryan)

- (a) *Ravii (ek) gaay k<sup>h</sup>ariidnaa caahtaa hai*  
 Ravi.NOM one cow buy.NF wish.IMPFV.3SG be.PRS  
 "Raw wishes to buy a cow (no particular cow in mind)"
- (b) *Ravii (ek) gaay-ko k<sup>h</sup>ariidnaa caahtaa hai*  
 Ravi.NOM one cow-DOM buy.NF wish.IMPFV.3SG be.PRS  
 "Raw wishes to buy a particular cow" (Mohanana 1994: 80)

Summing up, DOM in Hindi, as well as in other Indo-Aryan languages, obligatorily shows up when the direct object is human. Animate and inanimate direct objects get case-marked only if the speaker wants to construe them as definite and/or specific. As noted by Butt and King (1998) and Dalrymple and Nikolaeva (2011), the presence of DOM with a definite inanimate direct object strongly correlates with its topicality. This agrees with our prediction that, even in languages where DOM is no longer primarily conditioned by information-structure, topicality is the decisive factor that enables the extension and grammaticalisation of DOM to further NP classes.

Let us now turn to Dravidian languages. Two clear examples of DOM governed by animacy are Malayalam (Asher and Kumari 1997) and Kannada (Lidz 2006). DOM in Malayalam is mandatory with human and animate direct objects, as well as direct objects of worship, as in (5.30a, b, and c). According to Asher and Kumari (1997: 204) inanimate direct objects, which usually do not bear any case marking, may be overtly coded to resolve potential ambiguity, as in (5.30d), since word order is not a reliable clue in Malayalam:

(5.30) Malayalam (Dravidian, Southern Dravidian)

- (a) *avan kuttii-e aticcu*  
 he child-DOM beat:PST  
 "He beat the child"

- (b) *avan oru pafuvin-e vaayji*  
 he a cow-DOM buy:PST  
 "He bought a cow"
- (c) *aval filpatt-e araadhiccu*  
 she statue-DOM worship:PST  
 "She worshipped the statue"
- (d) *tiramaala-kal kappalin-e bheediccu*  
 wave-PL ship-DOM split:PST  
 "The waves split the ship" (Asher and Kumari 1997: 203)

Kannada DOM system seems to be closer to the systems found in other languages spoken in the Indian sub-continent. Human and animate direct objects are obligatorily case-marked, as in (5.31a, b). Inanimate direct objects may be optionally overtly coded. Lidz (2006) contends that, while overtly coded animate direct objects can be interpreted either as specific or non-specific, a specific interpretation is forced by the presence of the accusative marker on inanimate direct objects. In (5.31c), thus, the direct object "pustaka" must be interpreted as specific when it carries DOM. Conversely, the absence of the case marker on an inanimate direct object leaves room for both meanings:

(5.31) Kannada (Dravidian, Tamil-Kannada)

- (a) *naanu sekretari-yannu huDuk-utt-idd-eene*  
 1SG.NOM secretary-DOM look.for-NPST-be-1SG  
 "I am looking for a secretary" (Lidz 2006: 11)
- (b) *na-nu avan-annu du-dide*  
 1SG.NOM 3SG-DOM push-PST.1SG  
 "I pushed him" (Bhat 1991: 34)
- (c) *naa-nu pustaka-(vannu) huDuk-utt-idd-eene*  
 1SG.NOM book-DOM look.for-NPST-be-1SG  
 "I am looking for a book" (Lidz 2006: 11)

Interestingly, Lidz (2006: 12) also observes that the overtly coded inanimate direct objects are outside the scope of negation. In (5.32a), the interpretation would be either that I did not read any books or that I did not read a particular book. Crucially, with the overtly coded direct object, only the latter interpretation, with the direct object outside of the scope of negation, is possible, as in (5.32b):

(5.32) Kannada (Dravidian, Southern Dravidian)

- (a) *naanu pustaka ood-al-illa*  
 1SG.NOM book read-INF-NEG  
 "I didn't read a book"
- (b) *naanu pustaka-vannu ood-al-illa*  
 1SG.NOM book-DOM read-INF-NEG  
 "I didn't read a book"

Thus, DOM in Kannada, similarly to Hindi and many other languages, shows topicality effects in a way similar to Hindi. In fact, DOM with inanimate direct object correlates with properties typical of topics, such as being outside the scope of negation and having a definite or specific interpretation.

Likewise, DOM systems where animacy is the primary parameter governing DOM are found in Tamil (Lehmann 1993: 27-28; Schiffman 1999: 36-37) and Betta Kurumba (Coelho 2003). Human and animate direct objects are obligatorily overtly coded, whereas non-animate direct objects are optionally overtly coded only if they are definite and/or specific, like in Hindi and Kannada.

## 5.6 Creoles

Creoles are particularly interesting with regard to their DOM systems, since they often provide important evidence for the role of synchronic principles and the diachronic processes leading to the development and the regularities of morphosyntactic constructions. I will deal here with three Portuguese-based creoles (Diu Indo-Portuguese, South African Creole Portuguese, and Kristang) and one Malay-based creole (Manadonese). Although they are based on the same language, from which they presumably borrowed the constructions, Portuguese-based creoles differ significantly as to the use of DOM.

Diu Indo-Portuguese (Cardoso 2009) makes use of the prepositions *a* and *pə* to encode both dative arguments and DOM. The general distribution of these two forms is quite straightforward, even though some exceptions are found. Basically, *a* is used with pronominal arguments, whereas *pə* is found with every other NP (Cardoso 2009: 181). DOM in Diu Indo-Portuguese is primarily found with animate arguments, as shown by examples (5.33a, b, and c). In this case too, it is possible to overtly code inanimate direct objects optionally, presumably when they are definite or known to the hearer, as in (5.33d, e):

(5.33) Diu Indo-Portuguese (Creole, Portuguese-based)

- (a) *ikəl lion vey i rasp-o pə gat*  
 DEM lion come.PST and scratch-PST DOM cat  
 "The lion came and scratched the cat"

- (b) *el mem ater-o a el*  
 3SG EMPH push-PST DOM 3SG  
 "He pushed him"
- (c) *leopard foy murd-e pə lion*  
 leopard go.PST bite-INF DOM lion  
 "The leopard went and bit the lion"
- (d) *vay ve pə leyt*  
 go.NPST see-INF DOM milk  
 "Go check on the milk"
- (e) *vēt apəg-o pə vəl*  
 wind blow\_out-PST DOM candle  
 "The wind blew out the candle" (Cardoso 2009: 193 ff)

A similar situation is attested in other Portuguese-based creoles, such as South African Creole Portuguese, where the use of the marker *pro* is certainly attested with personal pronouns, as in (5.34):

(5.34) South African Creole Portuguese (Creole, Portuguese-based)

- ne misti dali pro mi!*  
 NEG must beat DOM 1SG  
 "Don't beat me" (den Besten 2000: 955)

In Kristang, DOM, expressed via the preposition *ku*, is restricted to animate direct objects. It is obligatory with pronouns and proper names, while it is optional with kinship terms, human and animate nouns, as shown by examples (5.35a, b, and c). The marker cannot be used with inanimate nouns (Baxter 1988: 150 ff; 1994):

(5.35) Kristang (Creole, Portuguese-based)

- (a) *eli ja dali ku yo*  
 3SG PFV hit DOM 1SG  
 "He/she hit me"
- (b) *yo ja olá ku Maria sa pai*  
 1SG PFV see DOM Maria POSS father  
 "I saw Maria's father"
- (c) *yo sa kanyóng ja olá (ku) ake femi*  
 1SG POSS elder.brother PFV see DOM that girl  
 "My elder brother saw that girl" (Baxter 1994)

Manadonese, a Malay-based creole spoken in Sulawesi, displays a DOM system similar to Kristang, based on animacy. The marker *pa* is compulsory with pronouns and proper names, while it is optional with kinship terms and animate nouns, as shown by the examples in (5.36). DOM is not allowed with inanimate nouns:

(5.36) Manadonese (Creole, Malay-based)

- (a) *Utu da skop pa kita*  
 Utu PST kick DOM 1SG  
 "Utu kicked me"
- (b) *Kita da tampeleng pa John*  
 1SG PST slap DOM John  
 "I slapped John"
- (c) *Dorang da kuti (pa) itu anak*  
 3PL PST flick DOM the child  
 "They flicked the child" (Wantalangi 1993)

DOM systems in all the creoles we have surveyed so far are straightforwardly based on the animacy of the direct objects. An intriguing feature of these creoles is the source morphology of their object markers. Some of them (like Diu Indo-Portuguese and South African Creole Portuguese) utilise prepositions derived from the Portuguese benefactive preposition *para*. This development is attested in another Romance language, namely Rumanian, which, unlike the vast majority of Romance languages, employs *pe* as object marker. Other creoles, such as Kristang and Bazaar Malay, use an adposition derived from a comitative marker, such as Kristang *cu*. This unusual development is well attested in some Portuguese- and Malay-based creoles spoken in South-East Asia (Baxter 1988). In addition, it is also found in the Hokkien dialects of Singapore and Malaysia, as shown by the examples in (5.37):

(5.37) Singapore Hokkien (Sinitic, Chinese)

- (a) *Gua kap i khua*  
 1SG DOM 3SG look  
 "I saw him"
- (b) *Gua yong chha kap i phha*  
 1SG use stick COM 3SG hit  
 "I hit him with a stick" (Baxter 1988: 168)



According to Baxter, since there is no DOM in Hokkien, this kind of system is likely to have been borrowed by Portuguese-based creoles. Nonetheless, the use of a comitative marker for the accusative function is independently well attested in many Sinitic languages, as Chappell (2006, in press) has demonstrated. In point of fact, this path is found in Min and Hakka dialects, as well as in some Southern-Mandarin dialects (see chapter 7). Based on Chappell's (2000, 2006, to appear), reconstruction, all these comitative markers can be traced back to verbal sources with the meaning "to accompany, to mix, to be the same as". These verbs later grammaticalised into comitative markers and coordinative conjunctions (a fact attested also in many West African languages). Afterwards, the comitative markers grammaticalised into an accusative marker, with an intermediate stage as a marker of benefactive and ablative case roles.

Thus, it would seem to be very likely that the unusual polysemy is not ascribable to the Portuguese superstratum. Rather, it could well be that the presence of this uncommon pathway of grammaticalisation among Portuguese- and Malay-based creoles is due to Sinitic influence. This issue, however, cannot be settled here.

## 5.7 Papua New-Guinea

The following languages of Papua New Guinea display either DOI or DOM systems based on animacy: eight of them are Trans New Guinea languages, i.e. Anamuxra, Dani, Tauya, Daga, Barai, Teiwa, Abui and Oksapmin, two are Sepik languages, Awtuw and Yessan Mayo. The last two, Imonda and Waris, belong to the Border family.

Anamuxra is a verb final language with no fixed word order (Ingram 2001: 310). It has an elaborated system comprising both animacy-based DOM and DOI; DOM is coded by the two case endings *-x* for the singular and *-xi* for the non-singular (i.e. dual and plural forms, Ingram 2001: 188), and there is DOI on the verb indicating the person and number of the direct object. DOM on personal pronouns is mandatory, as in (5.38 a, b, and c). It is instead optional with animate full NPs, as exemplified by (5.38d). DOI is always present with animate nouns, regardless of the kind of NP:

(5.38) Anamuxra (Trans New Guinea, Madang)

- (a) *ad-aku-m*            *ya-x*            *ya-tuwu-mna-ba-t*  
 DEM-FD-CLF.M 1SG-DOM.SG 1SG.OBJ-hit-IMM-FUT-3SG.SUBJ  
 "This (man) is about to kill me"
- (b) *Andrew-x*            *wara-pa-n*            *n-tamang-pa-ŋ*  
 Andrew-DOM.SG ant-CLF.RES-PL 3SG.OBJ-bite-FP-3PL.SUBJ

"Ants bit Andrew"

- (c) *nan-xi*                      *nag-kixr-i-n*  
 2PL-DOM.NSG 2SG-see-NT-1SG.SUBJ  
 "I saw you"
- (d) *Daniel-(x)*                *n-kixra-m*  
 Daniel-DOM.SG 3SG.OBJ-see-1SG.SUBJ  
*n-xiswar-i-n*                      *Andrew-(x)-pu*                *n-xiswara*  
 3SG.OBJ-tell-NT-1SG.SUBJ Andrew-DOM.SG-first 3SG.OBJ-tell  
 "I saw Daniel and told him "You go to talk to Andrew first"" (Ingram  
 2001: 188-189)

Interestingly, Ingram (2001: 308-309) notes that DOM and DOI behave differently in ditransitive constructions. While indexation obligatorily indexes the indirect object and never the direct object, both objects can be case-marked provided that the direct object (i.e. the theme) is animate as well. Unfortunately, no examples are provided that show case marking on both arguments in ditransitive clauses. A similar situation is attested in Oksapmin (Loughnane 2009: 156-160), where DOM is restricted to pronominal and human direct objects.

Daga has only DOI with human direct objects, and occasionally animate nouns, as shown by examples (5.39). Interestingly, there is no overt suffix for third person singular direct object (Murane 1974: 44):

(5.39) Daga (Trans New Guinea, Southeast Papuan)

- (a) *yawa-nege-n*  
 see-1SG.OBJ-3SG.PST  
 "He saw me"
- (b) *yawa-ge-n*  
 see-2SG.OBJ-3SG.PST  
 "He saw you"
- (c) *yawa-ne-n*  
 see-1PL.OBJ-3SG.PST  
 "He saw us" (Murane 1974: 44)

Dani exhibits a very interesting DOI pattern. In this language, only human (and, apparently, sometimes animate) direct objects can be affixed to the verb. The same affixes are used also for indirect objects and beneficiaries (Bromley 1981: 156). What is interesting about Dani is the fact that, as many other Trans New Guinea languages, it allows the object prefixes to be directly attached to the main verb only with a restricted subset of verbs. Otherwise direct object affixes

must be attached to post-cliticised auxiliaries that are usually compounded with the main verb stem, forming thus a sort of serial verb construction (see also Foley 1986). These auxiliaries are employed to signal semantic differences as well (see Foley 2000 for further information). As a matter of fact, object affixes can be attached only to the main verbs (*w*)*at-* "hit, kill", *hei-* "put", and *ha-* "perceive", as in (5.40a,b), whereas all of the other verb roots need an auxiliary-like verb such as *-hei-* "put", *-ha-* "see", *-et-* "give" or *-ap-* "do for" (the last two are used to introduce recipients and beneficiaries) to express the object affix overtly, as in (5.40c, d), Bromley 1981: 157-158):

(5.40) Dani (Trans New Guinea, West Dani)

- (a) *n-at-h-e*  
1 SG.OBJ-hit-REAL-3SG.SUBJ  
"He hit me"
- (b) *h-aheik-h-e*  
2SG.OBJ-put-REAL-3SG.SUBJ  
"They put you"
- (c) *ane-pu a'-la kela'-nakeikhe*  
its.noise-woo inside-it inward-3SG.SUBJ.put.1SG.OBJ  
"He put me on the airplane"
- (d) *hat joma hakako*  
2SG here 1PL.SUBJ.put.2SG.OBJ  
"We have stationed you here" (Bromley 1981: 156 ff.)

This feature, which seems to be unique to Papuan languages, needs to be investigated more thoroughly. While it seems proved that verb serialisation is the source of these constructions (Foley 1986: 142), to my knowledge, no explanations have been provided so far why the verbs meaning "kill", "perceive", and "put", which are quite heterogeneous from a semantic point of view, can freely host the object affix. In other languages, these auxiliaries have evolved into a transitivising strategy, as noted by Foley (2000: 378). Indeed, in Tauya (MacDonald 1993: 180), which has DOI for human direct objects, the auxiliary *-fe-* (which serves also to encode perfective aspect) has become a transitive suffix which is attached to the direct object prefix, as in (5.41):

(5.41) Tauya (Trans New-Guinea, Madang)

- (a) *?umu-nen-fe-i-?a*  
die-3PL.OBJ-TR-3PL-IND  
"They killed them"

- (b) *fanu moʔotu ʔumu-nen-fe-i-ʔa*  
 man many die-3PL.OBJ-TR-3PL-IND  
 "They killed many men"
- (c) *nono surifo ʔutine-fei-fe-a-ʔa*  
 child newborn fall-3SG.OBJ-TR-3SG-IND  
 "She dropped the baby" (MacDonald 1993: 180-181)

In another Trans-New Guinea language, Una (Louwrese 1988) the grammaticalisation process has gone even further, as the object prefix and the ancient auxiliary have become a single morpheme, which has been reanalysed as the object affix. In this language too, the presence of the object prefix is confined to human direct objects (Foley 2000: 378):

(5.42) Una (Trans New Guinea, Mek)

- se-kwan-ki-n*  
 cut-FUT-2SG.OBJ-1SG.SUBJ  
 "I will cut you" (Louwrese 1988)

Interestingly, Barai (Olson 1981) exhibits indexation only with human direct objects. Neither intransitive subjects nor agents are indexed on the verb, as shown by examples (5.43):

(5.43) Barai (Trans New-Guinea, Southeast Papuan)

- (a) *na ruo-vo*  
 1SG come-PRES  
 "I am coming" (Olson 1981: 13)
- (b) *fu na kan-ie*  
 3SG 1SG strike-1SG.OBJ  
 "He struck me"
- (c) *e ije bu na kan-ie*  
 man DEF 3PL 1SG strike-1SG.OBJ  
 "The people, (they) hit me" (Olson 1981: 24)

Similarly, Teiwa and Abui do not index either the agent or the subject (Klamer and Kratochvíl 2006). In Teiwa (Klamer 2010), only animate direct objects can be prefixed to the verb via the affix *g(a)*. Inanimate direct objects appear as independent pronouns, but are never indexed on the verb, as shown by (5.44a), where only

the animate direct object "us" is indexed on the verb. Examples (5.44 b, c) illustrate an interesting opposition. When the direct object is inanimate, as in (5.44b) and is expressed by a free pronoun, the verb means "to take something". When the direct object is animate, as in (5.44c), it is indexed on the verb and the verb means "to follow someone":

(5.44) Teiwa (Trans New-Guinea, Alor-Pantar)

(a) *a pi-liin muxui na*  
 3SG 1PL.INCL-invite banana eat  
 "He invited us to eat bananas"

(b) *na ga'an mar*  
 1SG 3SG take  
 "I take/get it" (Klamer 2010: 173-174)

(c) *na ga-mar*  
 1SG 3SG-follow  
 "I follow him"

As noted by Klamer and Kratochvíl (2006), even in the class of transitive verbs that always prefix direct objects, the prefixes for animate and inanimate direct objects are always distinct: a CVC prefix is used for animate objects, while a CV one for inanimates.

Abui (Kratochvíl 2007, to appear) shows a more complex picture. First, unlike Teiwa, Abui has a semantically based alignment system. Undergoers (i.e. subjects of stative intransitive verbs and patients) are coded by prefixes, unlike actors (i.e. agents of transitive verbs and subjects of intransitive motion verbs). Here I will set aside the discussion about the single argument of stative intransitive verbs. Even though I am using here the label "direct object" to refer to the second argument of a transitive predicate, it should be kept in mind that Abui does not provide any syntactic evidence for grammatical relations like subject and direct object (Kratochvíl 2007, personal communication). Nonetheless, for the sake of convenience, I will keep using this label here.

Some verbs, that typically take inanimate direct objects, do not allow an object prefix. Verbs that can have either an animate or inanimate direct object encode the inanimate/animate distinction by choosing among three different prefixes, summarised in table (5.7):

Affected DO	Non-Affected	Animate ± Affected
PAT	LOC	REC

Table 5.2: Abui object prefixes

The first set (PAT) is used for direct objects (either animate or inanimate) affected by the action expressed by the predicate. The second (LOC) and the third set (REC) encode direct objects that do not undergo a change of state or location, with the third set (REC) being restricted only to animate entities. For instance, the verb *fanga* "say" can combine with all three prefixes:

(5.45) Abui (Trans New-Guinea, Alor-Pantar)

- (a) *he-maama Simon ha-fanga*  
 3SG.AL-father Simon 3SG.PAT-say  
 "His father ordered Simon"
- (b) *ama he kang he-fanga*  
 person 3SG.OTHER can 3SG.OTHER-say  
 "People agree/approve (lit. Persons say it can)"
- (c) *a neng loku ho-fanga*  
 2SG man PL 3SG.HUM-say  
 "You scold at the men" (Klamer and Kratochvíl 2007)

Abui is particularly interesting because the use of the three set of prefixes is related to the affectedness of the direct object on the one hand and its animacy on the other one. Abui, along with Pomoan languages (see below, section 5.10), are the only examples of languages in which affectedness does play a discernible role in determining the presence and the distribution of DOM or DOI, contrary to what has been proposed by Næss (2004, 2007). Crucially, all these languages show a semantically aligned system, in which notions like control and affectedness are of paramount importance to account for the variation in the encoding of arguments. By contrast, affectedness turns out not to be relevant to account for the variation in direct object encoding in accusative languages, which indeed constitute the vast majority of the languages with DOM and DOI.

In Imonda, DOM is coded by the case ending *-m*, which indicates a variety of different semantic relationships (Seiler 1985), such as recipient, benefactive, direction, etc. Seiler (1985) states that DOM is obligatory when the risk of ambiguity occurs. This ambiguity is more likely to arise when both the agent and the direct object are human, as in (5.46a, b). However, DOM is found also with animates and can optionally occur with inanimates, as shown by examples (5.46c, d, e) respectively:

(5.46) Imonda (Border, Waris)

- (a) *aia-l edel-m ue-ne-uōl fe-f*  
 father-NOM human-DOM CLF-eat-PL do-PRS  
 "Her father habitually eats humans"

- (b) *mo-l-m ka-m f-ai-h-u*  
 daughter-NOM-DOM 1SG-DAT CLF-give-REC-IMP  
 "Give me your daughter!"
- (c) *tinbi ha-m ue-ne-fan*  
 python snake-DOM CLF-eat-PER  
 "The python has swallowed the (other) snake (*ha*=specific, small snake)"
- (d) *malhu-m ka falifiha-n*  
 pig-DOM 1SG shoot\_and\_miss-PST  
 "I had a shot at that pig but missed it"
- (e) *mëna-m kai nagla-i-me*  
 road-DOM Q see-PST-Q  
 "Did you see the road?" (Seiler 1985: 163-165)

These examples do not seem to be amenable to an explanation in terms of disambiguation only. As a matter of fact, DOM is not confined to situations in which there would be ambiguity, as shown by (5.46d, e). In these two latter examples, the agent is higher in animacy than the direct object, and extra-linguistic knowledge alone is sufficient to ensure the correct interpretation of the arguments, as it would be impossible to construe the pig in (5.46d) and the road in (5.46e) as agents of predicates like "shoot" and "see". In addition, as shown by (5.46b), the case marker *-m* encodes both arguments of a ditransitive construction, a fact that has been usually taken as increasing the risk of ambiguity.

Let us now turn to Waris (Brown 1981, 1988). Basically, DOM in Waris is found with animate direct objects, as shown by (5.47a, b). Interestingly, DOM is found on inanimate direct object as well, when the direct object is not affected by the action expressed by the verb, as exemplified by (5.47c), in which the direct object is completely affected and then not overtly coded, as opposed to (5.47d), where the direct object is not affected and thus gets case-marking, in a pattern that resembles the conative alternation found in English:

(5.47) Waris (Border, Waris)

- (a) *etel-va boasalel-m won-pró-i*  
 older.sibling-TOP young.sibling-DOM ACC:PL-come-PST  
 "The older brother just brought his younger brothers" (Brown 1988: 67)
- (b) *ka-va ye-m hlle-v*  
 1SG-TOP 2SG-DOM hear-PRS  
 "I am listening to you"

- (c) *ka-va ti he-v*  
 1SG-TOP tree chop-PRS  
 "I am chopping down a tree"
- (d) *ka-va ti-m he-the-v*  
 1SG-TOP tree-DOM chop-PRO-PRS  
 "I am chopping on a tree" (Brown 1988)

The last two Papuan languages, Awtuw (Feldman 1983) and Yessan Mayo (Foreman 1974), belong to the Sepik family. Awtuw overtly codes human direct objects with the suffixes *-re* and *-te* (the latter used only for feminine referents), which is also used to encode indirect objects and benefactives. Pronouns (including the demonstrative and interrogative ones), as well as proper names, take DOM obligatorily, as in (5.48a, b; Feldman 1983: 178 ff). Feldman also states also that DOM is mandatory when the direct object is equal to or higher than the agent in animacy, as shown by (5.48c). If the direct object in (5.48c) did not bear any case marking, the mandatory interpretation would be "The woman bit the pig". DOM seems to occur sporadically with inanimate as well as animate referents, as long as the direct object is highly definite, as in (5.48d):

(5.48) Awtuw (Sepik, Ramu)

- (a) *wan rey-e du-k-puy-ey*  
 1SG 3SG.M-DOM FAC-IMPV-hit-IMPV  
 "I'm hitting him"
- (b) *rey piyren Kampo-re d-ael-i*  
 3SG.M dog Kampo-DOM FAC-bite-PST  
 "The dog bit Kampo"
- (c) *Tey tale-re yaw d-ael-i*  
 3SG.F woman-DOM pig FAC-bite-PST  
 "The pig bit the woman"
- (d) *nemet rey tapwo-uyk-re d-ayn-e*  
 mother 3SG.M fire-smell-DOM FAC-smell-PST  
 "His mother smelt the odour of fire" (Feldman 1983: 182-183)

That the need for disambiguation is primary, nonetheless, is undermined by the possibility for both the direct and the indirect object to be overtly coded in ditransitive constructions, as in (5.49), as well as the presence of DOM in examples like (5.48d), where the direct object is an inanimate entity. Once again, if the major function of DOM were that of keeping the direct object distinct from



the agent, then we would expect that only one of the two objects in ditransitive constructions were overtly coded, since the presence of same case marking would increase the risk of ambiguity:

(5.49) Awtuw (Sepik, Ramu)

*yapo-re wan Kaempiy-te də-kow-o*  
 man-DAT 1SG Kaempiy-DOM FAC-give-PST  
 "I gave Kaempiy to a man " (Feldman 1983: 183)

In Yessan Mayo (Foreman 1974), human direct objects must be overtly coded by the suffix *-ni* (which encodes indirect objects and beneficiaries as well), as in (5.50a, b). The agent must precede the direct object in transitive constructions. This constraint alone would suffice to guarantee the correct interpretation of the role of the arguments. However, it is possible to find direct objects in initial position. In this case, the direct object bears case marking, as shown by (5.50c):

(5.50) Yessan Mayo (Sepik, Tama)

- (a) *an ti-ni kwotana*  
 1SG 3SG.F-DOM wait.3SG.PRS  
 "I'm waiting for her"
- (b) *an ti-ni aki-ye*  
 1SG 3SG.F-DOM fear.PST  
 "I was afraid of her"
- (c) *nim-ni yabel me Swagap-ki pim*  
 1PL-DOM sun tree Swagap-LOC fight/hit.3sg.pst  
 "It became noon when we were at Swagap (lit. The sun hit us at Swagap)" (Foreman 1974: 106 ff)

## 5.8 Tibeto-Burman

As is well known, many Tibeto-Burman languages display a DOM system primarily based on animacy (La Polla 1992, 1994, among others). While the cut-off points on the animacy hierarchy may vary considerably, and in some languages inanimate direct objects may be overtly coded as well, there is enough evidence to claim that animacy takes priority over other parameters in determining the distribution of DOM, at least from a synchronic point of view. From a diachronic point of view, the synchronic variation found within some Tibeto-Burman languages, as

well as diachronic evidence, seems to point to topicality as the triggering parameter for DOM. That is, similarly to other languages we have seen in this chapter, DOM was initially regulated by information-structural parameters, and was later extended to direct objects showing topical features.

This is in contrast with La Polla's (1992) view, according to which the function of DOM (which he calls "anti-ergative") in Tibeto-Burman languages is to disambiguate the semantic role of the participant. When an animate argument, which is usually an agent, occupies instead the direct object role, there is need to signal this situation. In a similar vein, the use of ergative case in these languages serves to encode a somehow non-prototypical agent (La Polla 1992, 1994, 2004). As we will see shortly, the available evidence seems to contradict the idea that DOM in Tibeto-Burman languages works as a pure distinctive means.

Thulung Rai (Lahaussais 2002) is a split ergative language, with the first person (both singular and plural), second singular, singular honorific and dual being encoded for nominative/accusative and the third persons, all common NPs as well as the second person plural taking ergative/absolute encoding. Animate direct objects are coded via the case marker *-lai*, used also for indirect objects (probably a borrowing from Nepali, which has DOM), as shown by the examples in (5.51):

(5.51) Thulung Rai (Sino-Tibetan, Tibeto-Burman)

- (a) *gu-ka gana-lai jal-na*  
 3SG-ERG 2SG-DOM hit-3SG/2SG  
 "He hits you"
- (b) *gu-ka khlea-lai jal-y*  
 3SG-ERG dog-DOM hit-3SG/3SG  
 "He hits the dog"
- (c) *gu-ka a-lwak-lai suq-dy*  
 3SG-ERG POSS.1SG-younger\_brother-DOM bring\_down-3SG/3SG  
 "He brought my younger brother down" (Lahaussais 2002: 68 ff.)

As noted by Lahaussais (2002: 70), an analysis in terms of a distinguishing function of DOM in Thulung is not tenable for two reasons. First, in transitive clauses both arguments are encoded on the verb, and therefore the need for distinguishing between participants is highly diminished, since indexation allows to ensure the correct interpretation. Second, even when participants are both third person, in which case verbal suffixes do not differentiate between the two arguments, the ergative marker *-ka* makes it clear which participant is the agent, as in (5.51c).

In Kham (Watters 2002), DOM behaves similarly to Thulung. The marker *-lai* appears with animate direct objects, as shown by examples (5.52). Once again,

there is indexation with both the agent and the direct object in transitive clauses, making the need to keep participants distinct less urgent:

## (5.52) Kham (Sino-Tibetan, Tibeto-Burman)

- (a) *na: no:-lai na-rĩ:h-ke*  
1SG 3SG.M-DOM 1SG-see-PFV  
"I saw him"
- (b) *nĩ: na-lai nə-rĩ:h-na-ke*  
2SG 1SG-DOM 2SG-see-1SG-PFV  
"You saw me" (Watters 2002: 67)
- (c) *o-nəĩ-lai kuwani-ni hai-k-o*  
3SG-friend-DOM well-ABL pull\_out-PFV-3SG  
"He pulled his friend out of the well" (Watters 2002: 240)

Meithei (also called Manipuri, Bhat and Ningomba 1997, Chelliah 1997) has a DOM system fundamentally based on animacy, in which only human/animate and specific direct objects are coded by the enclitic particle *-pu* (Chelliah 1997: 109), as shown by examples (5.53a, b):

## (5.53) Meithei (Sino-Tibetan, Tibeto-Burman)

- (a) *ma-ne Tombə-pu Chawbə-ta takí*  
3SG-CNTR Tomba-DOM Chawba-LOC point-NHYP  
"Tomba was pointed out to Chaoba by him"
- (b) *əŋáŋ-si má-pu ín-í*  
child-PDET 3SG-DOM push-NHYP  
"The child pushed him"
- (c) *əy-pu-na Ram-na nuŋsi-lə-pə-ti*  
1SG-DOM-CNTR Ram-CNTR love-PRO-NOM-DLMT  
*phə-kə-təw-ni*  
good-POT-OBLG-COP  
"If Ram (not Chaoba) loved me (and not Sira), it would be good"  
(Chelliah 1997: 109 ff.)

The case marker can also be used with recipients, experiencers, locations and causees, thus showing the typical polysemy of the accusative case well attested cross-linguistically. Interestingly enough, the Meithei accusative marker is homophonous with an 'adversative marker' that is employed to signal "that the *-pu* marked noun phrase is ill fated in being acted upon or that the verb is unexpected, unanticipated, or unfortunate" (Chelliah 1997: 117):

## (5.54) Meithei (Sino-Tibetan, Tibeto-Burman)

- (a) əy-pu      hi      hon-pa      həy-tə-e-ne  
 1SG-ADVR boat row-NOM proficient-NEG-ASRT-SS  
 "(But unfortunately), I don't know how to row a boat"
- (b) əy-pu-nə-pu                      Sekmay      čət-u      háy-í  
 1SG-DOM-CNTR-ADVR Sekmay go-IMP say-NHYP  
 "(Too bad), he ordered me (not you) to go to Sekmay" (Chelliah  
 1997: 118)

As shown in example (5.54b), the marker *-pu* may occur twice on the same NP, once as an accusative marker and once as a pragmatic marker. As will be discussed in Chapter 6, the homophony between the accusative marker and the pragmatic marker has been taken, for instance by Næss (2003: 1203) and Chelliah (2009), as evidence of the prominent role of affectedness in determining DOM. In their view, the meaning of unexpectedness or adversativity would derive from the "acted-upon, typically nonvolitional nature of the patient role" (Chelliah 2009: 398). In Iemmolo (accepted), I showed that affectedness is not a relevant parameter to account for the distribution of DOM in Meithei, and I proposed a different grammaticalisation path to account for the development of the "adversative" function" of the accusative marker in Meithei, which does not resort to the notion of affectedness. Here I will summarise the main points of my argumentation.

Usually, adversatives and, more generally, expressions of contrast or unexpectedness signal an event or opinion that is in contrast with the content of previous discourse, since they introduce a different viewpoint on an action or event. Consequently, they convey new or rather different information with respect to previous sentences and speaker's expectations. They may therefore be thought of as devices used to shift or amend information.

Although historical data on the grammaticalisation cline of the accusative marker in Meithei are not available, we think that an explanation which takes into account the information-structural constraints on DOM is more plausible, and supported by cross-linguistic evidence. I hypothesise that the feature that made the grammaticalisation process from accusative to adversative marker start is the topic function often associated with overtly coded direct objects (see chapter 6). Through a metonymisation process, the case marker *-pu*, initially used to indicate contrastive topic or topic shifts, comes to be associated with datives and human/animate and definite/specific direct objects because of the frequent connection between these features and topicality. The pragmatic marker is therefore reanalysed as an accusative marker, because this particular syntactic meaning is inferred and then associated with that form (Hopper and Traugott 1993, Cristofaro 2010).

The grammaticalisation process has gone even further, introducing an additional meaning of unexpectedness due to subjectification (Traugott 1995), as proposed by Chelliah (2009). In other words, through another pragmatic inferential process, speakers come to associate the topic shift marker with a subjective meaning of unexpectedness or adversativity. Something new may be unlikely to be expected. Most probably, as observed in Lemmolo (accepted), the marker became reanalysable as a marker of adversative via a stage as an emphatic particle, a usage still attested in Meithei (Bhat and Ningomba 1997: 147). The development postulated for Meithei could be thus represented in (5.55):

↗ adversative/unexpectedness marker

(5.55) topic shift/contrast → emphatic

↘ dative marker → topical object marker

Manange (Hildebrandt 2004) has a DOM system similar to the one found in Meithei. In this case as well, an alleged influence of affectedness in determining DOM has been proposed in order to explain the use of the object marker. As shown by examples in (5.56 a, b, c), DOM in Manange is limited to animate direct objects:

(5.56) Manange (Sino-Tibetan, Tibeto-Burman)

(a) *Iden-tse 3kyA-ri 2prin-tsi*

Eden-ERG 2SG-DOM hit-PFV

"Eden hit you"

(b) *IŋA-tse 3kyA-ri 2katti 3sA-pa siki ItsA-pa*

1SG-ERG 2SG-DOM many nice-NOM food eat-NOM

*Ipin-le 3kya Inuŋ-pa 1A-tsi*

give-CONCESS 2SG thin-NOM become-PFV

"Although I fed you many good foods, you still became thin"

(c) *nyko-tse nkor-ko-ri 1chin-tsi*

DEF-ERG cat-DEF-DOM catch-PFV

"The dog caught the cat" (Hildebrandt 2004: 116)

The Manange object marker *-ri* exhibits an interesting polysemy as well. Besides encoding experiencers and direct objects, it is primarily used as a locative marker to indicate "direction towards, as well as a sense of general location or to encode spatial and temporal deixis" (Hildebrandt 2004: 112). Interestingly, the case marker *-ri* is homophonous with a clitic *-ri* whose primary function is to

encode indefiniteness or newly introduced referents. Whilst Hildebrandt (2004: 121) claims that these two functions are in fact unrelated, it could well be that the two forms might have derived from the same source.

Finally, in Magar (Grunow-Hårsta 2004), humanness takes priority over the other parameters. While animate and inanimate direct objects must be highly topical and definite to receive case marking (Grunow-Hårsta 2004: 80), as in (5.57c), human direct objects are always overtly coded, even if indefinite or non-specific, as shown by (5.57b). The topicality requirement for animate and inanimate direct objects suggests that DOM is spreading to other NP classes, while its grammaticalisation with human nouns is now complete.

(5.57) Magar (Sino-Tibetan, Tibeto-Burman)

- (a) *Ram-e Kumari-ke dathup-a*  
 Ram-ERG Kumari-DOM hit-PST  
 "Ram hit Kumari"
- (b) *A-la-n MΛdebeni-an Thakal-ni-ki-ke*  
 R.DEM-CIR-LOC Madubeni-LOC Thakali-female-PL-DOM  
*bagΛ-dis-ca ta te-ola-mΛn*  
 sweep\_away-DER-ATT REP say-HAB-NOM  
 "It is said that there, at Madubeni, Thakali women were swept away  
 (in the flood)"
- (c) *hose-ko-i rfi-ke cΛkho jat-le cΛkho jat pyak*  
 DDEM-PL-ERG goat-DOM purify do-IPFV purify do after  
*hose-ko-i bΛli yafi-le*  
 DDEM-PL-ERG sacrifice give-IPFV  
 "They purify the goat and when it is purified they sacrifice it"  
 (Grunow-Hårsta 2004: 80)

## 5.9 South America

Many Southern American languages display DOM systems. In my sample, however, only three show a genuine system primarily governed by the animacy of the direct objects, since in many languages DOM turns out to be predominantly controlled by information-structural factors, as will be argued in Chapter 6. Even in the languages examined here, topicality influences DOM in the so-called optional contexts, similarly to other languages we have investigated in this chapter.

The three sampled languages surveyed here, Awa Pit (Barbacoan), Hup (Maku), and Kwaza (isolate) encode direct objects by way of a case suffix or postposition.

Awa Pit (Curnow 1997) is an AVO language that overtly codes only human direct objects via the postposition *-ta*, which is also used for indirect objects and location/direction. When the direct object is not referential, it stays uncoded, even though it is human, as shown by the opposition between (5.58a) and (5.58b). Example (5.58d) is interesting because it shows that the accusative *-ta* can be used several times in a sentence. If a distinguishing function were at work here, we would not expect to find the case marker both on the direct object "Santos" and the nominalised direct object of the matrix verb, as well as on the afterthought "woman":

(5.58) Awa Pit (Barbacoan, Pasto)

- (a) *ashaŋpa tita-mtu-s*  
 woman search-IPFV-LOCUT  
 "I am looking for a woman (to marry)"
- (a) *ashaŋpa-ta pyan-na-na wat shi ki*  
 woman-DOM hit-INF-TOP good NEG be.NEG.(NONLOCUT)  
 "Hitting one's wife is not good" (Curnow 1997: 266)
- (c) *Santos-ta-na miza pyan-a-ma-t*  
 Santos-DOM-TOP almost hit-PL:SUBJ-COMP-PFV.PART  
 "They almost beat up Santos" (Curnow 1997: 73)
- (d) *Santos-ta pyaŋ-ta-mika-ta pyan-ta-w*  
 Santos-DOM kill-PFV.PART-NMLRZ.SG-DOM hit-PST-LOCUT.SUBJ  
*ashaŋpa-ta*  
 woman-DOM  
 "I hit the one who killed Santos, the woman" (Curnow 1997: 288)

As in Maithili, pronouns use a different case marker for the accusative. Indeed, direct object personal pronouns are coded by the clitic *wa* for first and second person, and *a* for third person and the interrogative singular personal pronoun. Plural pronouns are coded by two different clitics *miza* (for first and second person) and *tuza* (for third person) (Curnow 1997: 85). We are therefore in the presence of a system in which pronouns are encoded differently from full lexical NPs. Since personal pronouns have a different accusative case, there is no need to add the object marker *-ta* to them, being already distinct.

In Kwaza (van der Voort 2004) DOM is regulated by the animacy of the direct object, even though cases in which an inanimate direct object is overtly coded are attested. Kwaza uses the case suffix *-wã* to overtly code animate direct objects, as in (5.59):

(5.59) Kwaza (Isolate)

- (a) *wã so-'wã e'xyi-ta-ki*  
 bee 1SG-DOM sting-1SG.OBJ-DECL  
 "(A) wasp stung me"
- (b) *zjuwãu-wã tsa'si-da-ki*  
 João-DOM follow-1SG-DECL  
 "I went after João"
- (c) *jere'xwa 'kay-ki natau-'wã*  
 jaguar scratch-DEC Natal-DOM  
 "The jaguar scratched Natal"
- (d) *aha-'wã ãwỹi-'ra*  
 father-DOM see-IMP  
 "Go and see your father" (van der Voort 2004: 105 ff.)

van der Voort (2004: 107) argues that, although at first glance DOM in Kwaza could be deemed a distinguishing strategy, this does not seem to be the primary purpose of the use of the object marker, as Kwaza verbs obligatorily index subjects and, with some verbs, direct objects, as shown by example (5.59a). DOM is used also in causative constructions and with transitivised verbs, provided that the direct object is animate.

A certain amount of variation is found within the domain of animate nouns, since the use of DOM seems to be optional in many cases in which one would expect to find the direct object obligatorily overtly coded, as in (5.60a, b), where the direct objects are a proper name and a kinship term respectively:

(5.60) Kwaza (Isolate)

- (a) *tsũhũ-'du e'tay(-tjate-wã) a'sa-re*  
 what-for woman-3SG.POSS-DOM leave-INT  
 "What a pity he left his wife!"
- (b) *zjwãu.(wã) butxi-'nã-da-ki*  
 João-DOM free-FUT-1SG-DECL  
 "I'm going to free João"

Finally, it seems quite interesting that question words are to be obligatorily coded by the case marker *-wã* regardless of the animacy, as shown by examples (5.61a, b):

(5.61) Kwaza (Isolate)



- (a) *tsūhū-'ra'ti-wa jari'mã jã-'re*  
 what-FOC-DOM jarimã be-INT  
 "What does *jarimã* mean?"
- (b) *dilɛ-'wã oi'tsi-tsy-hỹ-'re*  
 who-DOM copulate-GER-NMLRZ-INT  
 "Who's he going to fuck?"

DOM in Hup (Epps 2008, 2009) is predominantly driven by animacy, definiteness and specificity being less central for the presence of DOM. Pronouns, proper names, kinship terms, demonstratives and human nouns are obligatorily coded through the object suffix *-ǎn* (which, once again, is used for recipients as well as beneficiaries), while DOM is optional for animals and impossible with inanimates, as in (5.62, Epps 2008):

(5.62) Hup (Maku)

- (a) *?am-ǎn ?ɔt-yó? tih ham-yi?-iy*  
 2SG-DOM cry-SEQ 3SG go-TEL-DYNAM  
 "After crying over you, he left"
- (b) *hid-nih ?intúg-taeh-ǎn ti?cik-nih nin-ip*  
 3PL-POSS mother's\_husband-son-DOM dislike-NEG 2PL-DEP  
 "You all didn't dislike their step-brother either"
- (c) *?ayũp-?ih-ǎn ?ah ky-y j'ũg-ǎn*  
 one-M-DOM 1SG see-DYNAM forest-LOC  
 "I saw a man in the forest" (Epps 2008: 146 ff)

It is interesting that proper names are always overtly coded, regardless of the animacy of the direct object. In (5.63), the direct object *hǎt* "alligator" must be obligatorily overtly coded, even though it refers to a canoe:

(5.63) Hup (Maku)

- hǎt-ǎn ?ah d'oh-óh*  
 alligator-DOM 1SG take-DECL  
 "I took Alligator" (Epps 2008: 149)

In the realm of human nouns, DOM is mandatory when the direct object is specific. That is, indefinite non-specific direct objects are not case-marked, as opposed to the indefinite but specific ones, which are obligatorily overtly coded, as shown by examples in (5.64a vs. b). As for DOM with animate nouns, Epps clearly shows that it is optional. Nonetheless, when the animal is one of the main characters in stories, it gets case-marking almost invariably (Epps 2005: 151-152), being thus influenced by information-structural considerations:

## (5.64) Hup (Maku)

- (a) *waʔ taehʔin túk-úy*  
 buzzard child.mother want-DYNAM  
 "Buzzard wants a wife"
- (b) *waʔ tih-taehʔin-ǎn túk-'uy*  
 buzzard 3SG-child.mother-DOM want-DYNAM  
 "Buzzard wants his wife"
- (c) *tǎh-ǎn-mah j'ám tih wón-máh-ah*  
 tapir-DOM-REP DST.PST 3SG follow-REP-DECL  
 "He followed the tapir, long ago, they say" (Epps 2005: 151-152)

I conclude this survey of Hup DOM-system with a discussion of two interesting interactions with the DOM system found in Hup. The first one is the occurrence of the object marker in case of possessor raising, in which the human (and some animate) possessor of a body part is object-marked, as shown by (5.65): the use of the object marker to encode possessors is found in other languages as well, e.g. Tariana and Persian (see chapter 6):

## (5.65) Hup (Maku)

- (a) *tih-dóʔ-van pǎt ʔáh j'id-íy*  
 3SG-child-DOM hair 1SG wash-DYNAM  
 "I wash the child's hair"
- (b) *ʔám-ǎn ʔáh yoʔmǎy yók-tán-áh*  
 2SG-DOM 1SG anus stab-FUT.CNTR-DYNAM  
 "I'll stab you in the anus"

The second phenomenon, which appears to constitute an *unicum* cross-linguistically, is the unusual result of the interaction between plural marking and DOM. Plural marking in Hup (coded by *d'əh*) is limited to animate referents, being thus consistent with the rules that govern DOM. Unlike DOM, though, plural marking is not ungrammatical with inanimate nouns. Its occurrence with inanimate entities, however, seems to be quite rare in natural discourse (Epps 2009: 93). The intersection between these two subsystems leads to a very peculiar outcome. Indeed, when plural-marked nouns are in direct object position, DOM becomes obligatory regardless of the animacy of the direct object, as shown by examples (5.66a, b). The combination of the plural and object markers yields the fused form *-n'ǎn*):

## (5.66) Hup (Maku)

- (a) *biʔ-n'ǎn-mah yup tih haʔ-ʔe-h*  
 rat-PL.DOM-REP that.ITG 3SG search\_with\_hands-PFV-DECL  
 "He searched out rats with his hands, it's said"
- (b) *ʔah cug'aet-n'ǎn puhut-'əh-hi-yiʔ-iy*  
 1SG leaf/paper-PL.DOM blow-send-descend-TEL-DYNAM  
 "I blew the papers down" (Epps 2009: 94)

As convincingly observed by Epps (2009: 96 ff.), this typological unusual pattern can be explained diachronically. Since the conditions for both DOM and plural marking largely overlap, being both based on animacy, over the time this led to the systematic association of plural marking and DOM (and not the other way round, as it is the presence of plural marking that requires the presence of DOM). The presence of DOM with inanimate plural noun is due to the fact that DOM rules (which allow only animate direct objects to be overtly coded) are overridden by number marking rules, according to which inanimate nouns might be overtly coded as well.

## 5.10 Central and North America

Straightforward instances of DOM are hard to come by in North America: the most important concentration of DOM in this macro-area is found in Central America, especially in Oto-Manguean languages. I will start by examining the distribution of DOM in two Oto-Manguean languages, Copala Trique (Hollenbach 1992, López Cruz 2008) and Yaipetec Chatino (Rasch 2002).

Copala Trique marks animate direct object with the preposition *man*<sup>32</sup> (used also to encode the dative) whose original meaning "body of" is no longer available to speakers (López Cruz 2008: 48). Animate direct objects are optionally overtly coded, as opposed to pronominal direct objects, which must be obligatorily overtly coded, as in (5.67). Reportedly, DOM may be optionally found with inanimate specific direct objects as well, albeit no examples are provided (López Cruz 2008: 46):

(5.67) Copala Trique (Oto-Manguean, Mixtecan)

- (a) *kano<sup>1</sup> zhoh<sup>3</sup> shianh<sup>13</sup> zhoh<sup>3</sup> man<sup>32</sup> zii<sup>5</sup>*  
 POT.grab 3SG.AML POT.bite 3SG.AML DOM 3SG.M  
*ne<sup>13</sup> yoh<sup>3</sup> ra<sup>4</sup> zhoh<sup>3</sup> a<sup>32</sup>*  
 CON2.sit there CON.think 3SG.AML DECL  
 "It would grab the person sitting there and would bite him, it thought"

- (b) *nehe*<sup>3</sup> *zoh*<sup>3</sup> *man*<sup>32</sup> *gwaa*<sup>4</sup> *a*<sup>32</sup>  
 CON.see 3SG.F DOM John DECL  
 "She sees John" (Hollenbach 1992: 189)
- (c) *C-uchruj*<sup>32</sup> *ra'nga*<sup>3</sup> *cha*<sup>3</sup> *na*<sup>1</sup> *man*<sup>32</sup> *unj*<sup>1</sup> *a*<sup>32</sup>  
 COMPL-put lazo woman DOM 1SG DECL  
 "The woman cursed me" (López Cruz 2008: 207)

An interesting fact about Copala Trique, which is at odds with both the Distinguishing and the Indexing hypotheses, is the occurrence of the object marker on reflexives, which are expressed by a possessive phrase containing *mahan*<sup>13</sup> (Hollenbach 1992: 188), as exemplified in (5.68):

(5.68) Copala Trique (Oto-Manguéan, Mixtecan)

- tikawik*<sup>3</sup> *zoh*<sup>3</sup> *man*<sup>32</sup> *mahan*<sup>13</sup> *zoh*<sup>3</sup> *a*<sup>32</sup>  
 COMPL.kill 3SG.M DOM REFL his DECL  
 "He killed himself" (Hollenbach 1992: 188)

The sentence in (5.68) represents a direct reflexive construction (Kemmer 1994), in which the single participant is linguistically construed as two distinct entities, which refer to the same entity in the real world. Since the event involves two participants from a linguistic point of view, the formal rules requiring the overt coding on human direct objects are therefore applied. Nevertheless, this pattern seems to be very marginal, and Copala Trique constitutes the only language that may overtly coded reflexives with DOM in the sample on which this study is based.

Yaitepec Chatino (Rasch 2002) overtly codes all pronominal direct objects, whether animate or not, via the adposition *7in*. Full NPs are overtly coded only if animate, the inanimates being left uncoded, as in (5.69):

(5.69) Yaitepec Chatino (Oto-Manguéan, Zapotecan)

- (a) *nw-sla* *nu* *kwchi-kwtzen* *t7wa* *ku*  $\emptyset$  *7in*  
 COMPL-open DET lion mouth POT.eat 3SG DOM  
 "The lion opened its mouth to eat him"
- (b) *y-ja7* *7wi* *tne* *s7en* *nty-ku* *kwna* *ka-ti* *ke*  
 COMPL-sleep be blood place HAB-eat snake seven head  
*7in* *ntten*  
 DOM person  
 "They slept where there was blood because the snake with seven head was eating people"

- (c) *n7a Liya 7in nu ki7yu kwa*  
 COMPL.see Maria DOM DET brother that  
 "Maria saw the brother" (Rash 2002: 207-216)

DOM with direct objects other than pronouns seems to be optional. The sentence (5.69c) would be indeed acceptable without the adposition, without any change in the semantics of the portrayed event. As Rash (2002: 217) has noted through the analysis of narrative texts, however, the presence of the object marker implies a higher degree of discourse saliency of the direct object. As will be shown in Chapter 6, this situation seems to be confirmed by the data of another Chatino variety, Zenzopetec, in which the presence of the marker is related to the information status of the direct object.

There are not many North American languages that exhibit either DOM or DOI in my sample. DOM systems in which the animacy of the direct object is the predominant parameter are found in Zuni (isolate), Mohawk (Iroquoian) and Eastern and Central Pomo (Hokan).

Zuni (Nichols 1997) is tendentially an SOV language, but word order can be changed for pragmatic reasons. First and second person pronouns are inflected for case, while there are not either third singular or plural pronouns. There is no case marking on nouns except for the accusative *-ya'* (used also for indirect objects and possessors), which is obligatorily found on proper nouns and optionally with human nouns, as in (5.70). Nichols (1997) argues that the usage of the accusative with human nouns is intended to convey definiteness, as shown by example (5.70c), in which the noun *okyat* "girl" is interpreted as definite due to the presence of the marker:

(5.70) Zuni (isolate)

- (a) *ho-m Nemme' 'ansattu-kya*  
 1SG-DOM Nemme help-PST  
 "Nemme helped me"
- (b) *ho' Nemme-ya' 'ansattu-kya*  
 1SG.NOM Nemme-DOM help-PST  
 "I helped Nemme"
- (c) *ho' okyat-ya' 'ansattu-kya*  
 1SG.NOM girl-DOM help-PST  
 "I helped the girl (not the boy)" (Nichols 1997: 19-20)
- (d) *aktsek'i-ya' watsit utde-kya*  
 boy-DOM dog bite-PST  
 "The dog bit the boy" (Cook 1974: 38)

It is worth noting that in (5.70d), the overtly coded direct objects are fronted. As a matter of fact, Nichols (1997: 21) states that an overtly coded direct object can be freely moved in the clause. There seems thus to be a correlation between the presence of the marker and information structure parameters. Unfortunately, the available data does not make it possible to investigate this correlation more thoroughly. Nichols (1997: 20) further claims that the accusative suffix *-ya'* is in complementary distribution with determiners and demonstratives, such as *lukkya* "this" and *ussi* "that". Although she does not provide any examples in real sentences, evidence for the likelihood of this pattern comes from Corsican (Indo-European, Romance), which has a DOM system based on humanness. In Corsican the use of the preposition *a* is prohibited when nominals are modified by determiners and quantifiers as in (5.71, see Neuburger and Stark submitted):

(5.71) Corsican (Indo-European, Romance)

*Vigu- (\*a) l'omu*  
see-1S.PRS DOM the-man

"I see the man" Neuburger and Stark (submitted)

Although more investigation is needed before some conclusions can be drawn, the unusual incompatibility of DOM and determiners found in Zuni and Corsican suggests that DOM has undergone a process of hyperanalysis (Croft 2000), by which the functional scope of the object marker is broadened to encode definiteness and, most importantly, pragmatic identifiability. Since the presence of DOM in both of these languages involves that the direct object is definite and identifiable, over time this led to the systematic association of DOM with the semantic content carried by determiners. As a result, an inference of definiteness and identifiability arose, and this triggered the complementary distribution between DOM and determiners found in Zuni and Corsican. Eventually, once this association has been established, the usage of both DOM and determiners at the same time became redundant, since they convey the same value.

Eastern and Central Pomo (Hokan, Pomoan) are semantically aligned languages (see Donohue and Wichmann 2008), in which the morpho-syntactic encoding of participants (especially the arguments of intransitive predicates) is regulated by the semantic role they have in the event. Similarly to the Abui case discussed above, subjects of intransitive predicates can be encoded as the agent of a transitive predicate or as the direct object, based on some semantic parameters, such as control and affectedness. In Central Pomo (Mithun 1991, 1999, 2008) as well as in Eastern Pomo, only pronouns and nouns referring to human beings can carry case markers (Mithun 1999: 217): the accusative case is used on direct objects with whom the speaker empathises, namely human beings, as in (5.72):

## (5.72) Central Pomo (Hokan, Pomoan)

- (a) *mú-tu*    *ʔa*            *hk-úm*  
 3SG-ACC 1SG.NOM kill  
 "I killed him"
- (b) *mu:l*    *ʔa*            *hk-úm*  
 3SG 1SG.PAT kill  
 "I killed it (a bee)"
- (c) *ʔa:*            *mú-tu*            *ʔé:čadiw*  
 1SG.NOM 3SG.ACC chase\_away  
 "I chased him away"
- (d) *mu:l*            *to:*            *ʔé:čadiw*  
 3SG.NOM 1SG.ACC chase\_away  
 "He chased me away" (Mithun 1999)

Similarly, in Eastern Pomo (McLendon 1975, 1978), pronouns are accusatively coded (first and second singular person via special morphology), while human nouns are coded by the suffix *-al* (McLendon 1975: 156; 178), as in (5.73):

## (5.73) Eastern Pomo (Hokan, Pomoan)

- (a) *xá:su:l*    *wí*            *ko:k<sup>h</sup>óya*  
 rattlesnake 1SG.ACC bite  
 "A rattlesnake bit me"
- (b) *há*            *mí.pal*            *šá:ka*  
 1SG.NOM 3SG.ACC kill  
 "I killed him" (McLendon 1978: 1; 7)
- (c) *ʔíqanxa,*    *mí.p*            *ma.mq-al*                            *šá.kle báya*  
 be.after.HYS 3SG.NOM POSS.3SG-older.brother-ACC kill there  
 "And then, he killed his older brother there" (McLendon 1978: 179)

In both languages, the presence of accusative case marking on direct objects seems to be related to the lack of control on the one hand and affectedness on the other one. As elegantly argued by Mithun (1991), the use of the accusative case in these languages is reserved to participants that do not control an event and are affected by the event itself (Mithun 1999: 218). At first glance, this pattern would appear to be in conflict with my claim that affectedness plays only a marginal role in DOM systems, if any. However, it should be noted that, in this case, it is the overall alignment system that is clearly dictated by semantic features like

affectedness. The choice of the accusative case on direct objects or intransitive subjects is related to the parameter of affectedness. Affectedness, however, is not relevant in nominative-accusative languages, in which the assignment of DOM is primarily influenced by information-structural features like topicality.

Likewise, Mohawk (Iroquoian, Baker 1996) displays an interesting pattern. In this language, animate direct objects are either coded via indexation on the verbal auxiliary or incorporated into the verb. When a direct object is incorporated into the verb, indexation is prohibited, and vice versa, as shown by the opposition between (5.74a) and (5.74b):

(5.74) Mohawk (Iroquoian, Northern-Iroquoian)

- (a) *Shako-núhwe'-s*                      *ne* (*owirá'a*)  
 SG.M.3PL.OBJ-like-HAB NE baby  
 "He likes them (babies)"
- (b) *Ra-wir-a-nuhwe'-s*  
 SG.M-baby-like-HAB  
 "He likes babies" (Baker 1996: 316)

As is well known, incorporation of animate arguments is strongly disfavoured, and only a few nouns can be incorporated, indexation being instead strongly favoured with animate nouns. Though Baker (1996: 316) states that the reasons why animate nouns resist incorporation are unclear from a Minimalist point of view, a suitable explanation has been provided by Mithun (1984). She shows that direct object incorporation often serves to "background an argument" (Mithun 1984: 863): animate direct objects are usually very important in discourse, and therefore they do not get incorporated. Indeed, they trigger indexation, which has been shown to be strongly connected with topicality (see Siewierska 2004: ch. 4). Mohawk seems to have grammaticalised this tendency. The complementary distribution of indexation and noun incorporation is readily explainable when this discourse function is taken into account. Animate direct objects tend to trigger indexation more often than incorporation because they are more topic-worthy than inanimate entities. Incorporation occurs in the (quite rare) case that the animate direct object is backgrounded. Other factors, however, enter into the likelihood for a direct object to be incorporated, such as the kind of predicate and the degree of referentiality of the noun (see Mithun 1984 for further discussion).

## 5.11 Africa

DOM and DOI are quite widespread in African languages. However, systems in which animacy takes priority over the other parameters are not very common



amongst the languages of Africa. As will be argued in Chapter 6, both DOM and DOI in many African languages appear to be sensitive to information structure features, such as topicality. Within the Niger-Congo phylum, an object marking system primarily regulated by animacy is found in Noon (Soukka 2000) and Makwa (Stucky 1981, Morimoto 2002). For the sake of convenience, Makwa will be discussed along with the other Bantu languages in Chapter 6.

In Noon, animate direct objects are obligatorily indexed on the verb as suffixes (Soukka 2000: 190). They normally occur on the finite verb of the predicate, even though in some cases DOI can be found on the infinitive, when multiple auxiliaries are present in the clause, as in (5.75c):

(5.75) Noon (Niger-Congo, Atlantic-Congo)

- (a) *Mi hay-raa ki-feek*  
 1SG AUX.FUT-2SG.OBJ INF-spank  
 "I'll spank you"
- (b) *Kodu lóm-íd-rii-ri dara*  
 Kodu buy-TR-NEG-3SG.OBJ nothing  
 "Kodu hasn't bought her anything"
- (c) *Mi san ki-jéem-ri ki-'ap*  
 1SG refuse.AUX INF-try-AUX-3SG.OBJ INF-kill  
 "I refuse to try to kill it" (Soukka 2000: 190)

Interestingly, in ditransitive constructions, the order of suffixes (as well as the order of full NPs) is fixed, with the suffix indicating the indirect object always preceding the one indexing the direct object, as in (5.76):

(5.76) Noon (Niger-Congo, Atlantic-Congo)

- Teew-aa-ríi-ri!*  
 show-IMP-1PL.EXCL.OBJ-3SG.OBJ  
 "Show him/her to us" (Soukka 2000: 190)

Lango is a Nilo-Saharan language spoken in Uganda, in which human pronominal direct objects are indexed on the verb via a suffix (Noonan 1992), as in (5.77a). If a benefactive verbal stem is used, only the benefactive can trigger indexation. On the contrary, in ditransitive constructions, though either direct objects or indirect objects may trigger indexation, indexation of direct objects is clearly favoured (Noonan 1992: 141 ff.), as shown by example (5.77b), in which the direct object is indexed on the verb and the indirect object is introduced by the preposition *bót*. Non-human singular direct objects are never indexed on the verb:

(5.77) Lango (Nilo-Saharan, Eastern Sudanic)

(a) *okele*

3SG.SUBJ.bring.PFV-1SG.OBJ

"He brought me"

(b) *rwot ocwale*

*bote*

king 3SG.send.PFV-3SG.OBJ to.1SG

"The king sent him to me"

## 5.12 Conclusion

In this chapter I have investigated the cross-linguistic behaviour of DOM and DOI with respect to animacy. The data presented so far confirm my hypothesis, proposed in Chapter 3, that, although both DOM and DOI are controlled by the same information-structural factors, they behave differently with respect to the function they perform in discourse.

Whilst both are driven by the very same discourse prominence features, the two constructions show fundamental differences as to their functions. DOM systems seem to be fundamentally a device to signal the unexpectedness of an animate referent in the direct object role through case marking. DOI systems, on the contrary, signal the high discourse prominence of these direct objects and serve as a reference-tracking device, as confirmed by the fact that they often occur without the NP they refer to (Barlow 1992).

An interesting fact that emerges from the data analysed in this chapter concerns the actual role of the discriminating function that has been traditionally suggested to account for the appearance and distribution of DOM.

Let us briefly consider what the data tells us about the use of DOM as an ambiguity avoidance strategy. As I have discussed in Chapter 3, the idea that DOM serves to distinguish agents and direct objects at the clause level seems to be reductive, since it does not take into consideration i) the fact that other clues, derived from the discourse context or world knowledge, are usually available to the speaker for the correct assignment of function within a clause, and ii) empirical data show that in fact DOM may occur even when there is no need for disambiguation. Indeed, the role of distinguishability is greatly undermined by the data presented in this chapter.

Consider for instance the cases of Nhanda (section 5.3), Imonda, Awtuw, and Anamuxra (section 5.7). In these languages, the differential object marker is identical to the dative marker, and the same case marker appears both on direct objects and indirect objects, provided that both are human and/or animate. If a differentiating function were at work in these cases (note that for Imonda, Seiler clearly

states that DOM serves to distinguish between agents and direct objects), then ditransitive constructions should not display the same double marking of their objects, since this hinders the correct interpretation of the roles within the clause, producing ambiguity rather than solving it. The same case marker, thus, is used to indicate participants that may potentially be confused, since they are both animate and/or human. In some cases, as in Nhandá, other clues seem to be available, such as the relative ordering of direct and indirect objects. Nonetheless, quite a few languages show this "increasing ambiguity" strategy, such as the double marking strategy (found, for instance, in Dolakha Newari, Retuarã, Udi, Korku, Kinyarwanda, to name only a few, see Chapter 6), or the zero marking strategy, where both objects show no marking at all. The latter strategy seems to be quite common cross-linguistically (see Haspelmath 2008 for discussion).

That the use of DOM goes beyond the scope of the clause and is influenced by the discourse context is made clear also by its occurrence in cases in which the need for disambiguation is null. Indeed, in many languages, DOM comes about even when there is no risk of ambiguity. For instance, in Awtuw, as well as in Imonda and Yessan Mayo, DOM has been claimed to carry out a distinguishing function. In all of these cases, though, DOM occurs when agent and direct object are different in animacy (Awtuw and Imonda), or when a word order constraint that requires the agent to precede the direct object renders superfluous the use of DOM (Yessan Mayo).

A similar analysis covers the case of the Tibeto-Burman languages sampled in this chapter, which have been taken, for instance by LaPolla (1992, 1994), as a case in which DOM purportedly distinguishes between agent and direct objects. As we have seen in section (5.8), however, all of the Tibeto-Burman languages that show DOM systems based on animacy have other clues that resolve the possible ambiguity between agents and direct objects, such as indexation (Thulung Rai, Kham) or the use of case marking on agents (Manange, Magar). The differentiation between the two arguments of a transitive clause, be they human and/or animate, is already carried out by other strategies, as in the case of indexation. Likewise, most of the languages surveyed in this chapter can be easily showed not to be influenced by the need for disambiguation. Indeed, the role of ambiguity has been over-emphasised, whilst little attention has been paid to the interaction of DOM with other strategies, such as indexation and case marking on agents, which greatly undermine the importance of ambiguity avoidance.

Amongst the languages analysed, the only cases in which distinguishability between participants seems to be the underlying factor that prompts DOM are Malayalam, Kashmiri, and Yongren Lolo discussed in Chapter 2 (see section 5.5). In the former case, it seems that DOM is used whenever ambiguity may occur, as exemplified by (5.30d), in which both arguments are inanimate. In the latter case,

instead, DOM appears whenever the direct object outranks the agent in animacy, as shown by the examples in (5.24).

What brings about the animacy-based DOM systems we have seen so far? In my view, animacy-based DOM systems, as well as DOI systems, are a reflection of the topicality of animate direct objects, which becomes conventionalised (through grammaticalisation) as a restriction to human and/or animate direct objects (see Chapter 4 for a discussion of the close connection between animacy and topicality). That is, in most of the languages analysed in this chapter, animacy became the predominant parameter over topicality, because animacy is a typical feature of topical referents.

There are plenty of such examples among the languages investigated in this chapter. Consider the case of Arabic Lebanese. In Arabic Lebanese, as well as in Kwaza, (section 5.2), for instance, DOM is regulated by animacy, to the extent that question words (which are prototypically focal elements) must be obligatorily overtly coded if they are human. This is due to the fact that animacy, over time, took priority over topicality, thus causing DOM to be extended to all the direct object showing topic-worthiness features, like animacy (see Chapter 8 for diachronic data from Old Sicilian). Thus, every direct object that fulfils the right semantic conditions is overtly coded, regardless of its information-structural status. Yet, the role of topicality is still observable in the so-called "optional" contexts, where the use of DOM becomes regulated by information structure.

Likewise, definiteness and specificity often affect the distribution of DOM. As we have seen in Chapter 4, both definiteness and specificity positively correlate with topicality. Indeed, many studies have shown that definite and specific NPs are strongly favoured as topics in discourse. In the languages examined in this chapter, many cases are found in which there is a further split in the marking of animate (as well as inanimate) direct objects based on definiteness and/or specificity distinctions. For instance, all the languages of the Indian subcontinent I have analysed (see section 5.5) show such a pattern, in which animate and inanimate direct objects are overtly coded if the direct object is construed as topical (definite and/or specific). A similar motivation applies to Tibeto-Burman languages (section 5.8).

The chief role of topicality in determining DOM and DOI is confirmed by the data that I will analyse in the next chapter, in which the primary trigger for DOM and DOI is the definiteness and/or specificity of the direct object. As we will see, in this case the chance of confusion between the agent and the direct object is even lower, since the assignment of the agent function to a particular NP is mainly driven by animacy. Another challenge to the Distinguishing Hypothesis, as well as to the Indexing one, comes from the very peculiar pattern found in Copala Trique (see 5.10), which uses DOM with reflexive objects, which contradicts the idea that i) DOM serves to distinguish between participant in the clause on the

one hand, or ii) that DOM signals a high degree of transitivity. A very tentative explanation of this anomaly lies in the grammaticalisation of the object marker with animate objects, which comes to be extended to every direct object provided that it is human. However, in the absence of further data as to the frequency of use of this construction, it is impossible to solve this question here.

To sum up, the data presented in this chapter confirms our idea that the main function of animacy-based DOM systems is not to distinguish between agent and direct objects (*contra* Aissen 2003, Malchukov 2008, de Hoop and Malchukov 2008). Indeed, in the languages surveyed in this chapter, a clear distinguishing function is at work in only two cases. In the next chapter, we will see that the vast majority of DOM and DOI systems are overwhelmingly associated with information structure properties, such as topicality.

## Chapter 6

# A typology of DOM and DOI: topicality and definiteness

### 6.1 Introduction

In this chapter, I will examine DOM and DOI systems in which topicality takes priority over the other parameters, i.e. languages in which the main parameter governing DOM and DOI is topicality. As we have seen in Chapter 5, in the majority of languages that have DOM and DOI, seldom is one parameter responsible for the appearance of the phenomena under investigation. In this chapter, I set out to demonstrate that DOM and DOI systems are means to express the topicality of the direct object. I will also show that, despite being both governed by topicality, DOM and DOI have different functions in discourse. While DOM serves to signal topic-discontinuities, such as topic-shifts and topic-promotions, DOI is associated with highly continuous topics, and functions as a reference-tracking device. This difference is supported by syntactic evidence. DOM shows a recurrent association with topic-marking constructions, such as dislocations, topicalisations and fronting. Indeed, as we will see shortly, in many languages DOM is closely related to the pragmatic structuring of the discourse, as it occurs (and often arises) in topic-shift and topic-promotion constructions, such as dislocation.

I will also show that animacy-based DOM and DOI systems, such as the ones discussed in the previous chapter, are the result of the grammaticalisation of the construction(s) from earlier topicality-based systems. This is also evident when the apparent optionality found with some NP classes in animacy-based DOM and DOI systems is taken into account. We will see that this optionality is indeed due to the differences in topicality, with more topical direct objects being more prone to take overt coding than non-topical ones.. In this chapter I will also discuss three languages in which DOM is regulated by the definiteness of the direct object,

without any strong influence of information-structural factors. The first one is Modern Hebrew, where every syntactically definite direct object must be overtly coded (see below, section 6.2). The second one is Malagasy, where the object marker is nowadays mandatory with pronouns and proper names, as well as with demonstratives and one of the two definite articles. The third one is Hungarian, where the presence of DOI is triggered by the formal definiteness of the direct object. These are the only cases in which definiteness in the main governing parameter.

As in the previous chapter, DOM and DOI systems will be analysed based on language families, wherever possible. Otherwise, macro-areas will be taken into consideration.

## 6.2 Afro-Asiatic

Many Afro-Asiatic languages display either DOM or DOI, or both (see Chapter 5). As we have seen in Chapter 5, in some Afro-Asiatic languages both DOM and DOI are primarily governed by the animacy of the direct object referent they code and/or index. In all the languages that I am analysing in this section, DOM is the sole available option, with the exception of the Aramaic dialect of Tekelpe, in which both strategies are available.

I will start now by examining the distribution of DOM in Biblical and Modern Hebrew. Since the usage and the development of the object marker *'et* have been minutely discussed in several publications, I will not dwell on many technical and philological details, due to limitations of space. Rather, I will outline the distribution of DOM, highlighting some interesting facts that are not well known to theoretical and general linguists. The general rule of thumb states that, in Modern Hebrew, the object marker *'et* occurs when the direct object is definite, as in (6.1a). If the direct object is indefinite, it stays uncoded, as in (6.1b):

### (6.1) Modern Hebrew (Afro-Asiatic, Semitic)

(a) *baláti*                      *et*      *ha-zvuv*  
 swallow:1SG.PST    DOM    DET-fly  
 "I swallowed the fly"

(b) *baláti*                      *zvuv*  
 swallow:1SG.PST    fly  
 "I swallowed a fly" (Glinert 1989: 157)

Basically, the object marker is obligatorily found when the direct object is inherently or overtly definite (e.g. personal pronouns, proper names, and direct

objects modified by a determiner or a possessive, see also Givón 2001: 470). The usage of *'et* in other contexts is however quite complex. As noted by Glinert (1989: 157 ff.), *'et* can be dropped in formal and literary contexts (see Danon 2002: 74). In addition, *'et* is found with the existential-locative particle *yeš* in possessive constructions, as shown by (6.2):

(6.2) Modern Hebrew (Afro-Asiatic, Semitic)

*yeš le-Dan 'et ha-zefer ha-ze*  
 exist to-Dan DOM the-book the-this  
 "Dan has this book" (Danon 2002: 78)

Danon (2002) analyses the DOM system of Modern Hebrew as representing a formal distinction between NPs and DPs. In his view, only direct objects that show syntactic definiteness (that is, those that trigger N-to-D raising) can get *'et*, while indefinite direct objects do not get overt coding because they do not carry syntactic definiteness, i.e. they are NPs as opposed to DPs. Thus, in Danon's view, the presence of *'et* is only a syntactic fact, due to the presence of a determiner that renders the direct object syntactically definite.

The use of DOM in Modern Hebrew constitutes an *unicum* cross-linguistically, since every definite object is overtly coded without exception. The obligatoriness of this rule could seem a little bit artificial, and it is probably a result of the way Modern Hebrew has been revitalised. If one looks at Biblical Hebrew, the situation seems to be much more complex. Biblical Hebrew has DOM with definite direct objects, as in (6.3a). DOM is customary with personal pronouns and proper names (Waltke and O'Connor 1990), while it is optional with other categories of definite NPs. Unlike Modern Hebrew, though, Biblical Hebrew allows the omission of the object marker, even when the direct object is definite, as in (6.3c), in which the direct object "libations" is left uncoded even if it is modified by the possessive, as opposed to (6.3b):

(6.3) Biblical Hebrew (Afro-Asiatic, Semitic)

- (a) *bə-re'šit bārā 'elohim 'et ha-ššamayim*  
 in.beginning create.3SG.M.PFV god.PL DOM DET-heaven.PL  
*wə 'et hā-'āres*  
 and DOM ART-earth  
 "In the beginning, God created the heavens and the earth" (Gen. 1:1;  
 cited in Farina 2006)
- (b) *vajasesx 'et nisk-o*  
 pour.3SG.M.PST DOM libation-his  
 "He poured his libation" (2 Kings 16:13; cited in Berry 2006)



- (c) *bal ʔasix niskej-hε m midam*  
 NEG pour.1SG libations-their of\_blood  
 "I will not pour out their blood libations" (Ps 16:4; cited in Berry 2006)

The difference between (6.3b) and (6.3c) might be also related to the presence of negation. As we will see below, in Sahidic Coptic, DOM is hardly found within negated clauses.

The range of variation found in Biblical Hebrew is rather complex. Apart from definiteness, other parameters closely interact with the presence of DOM. Indeed, as demonstrated by Bekins (in preparation) in his corpus study on DOM in Biblical Hebrew, DOM in so-called "optional" contexts is influenced by other parameters as well. Bekins found that the apparent optionality of DOM becomes clearer when animacy, identifiability, and topicality are taken into account. His findings show that human and animate direct objects are more likely to be overtly coded as opposed to inanimate direct objects, provided that they are definite (87% vs. 64%). The main factor that underlies the use of DOM is topicality (Bekins in preparation). Using Givón's measures of referential distance and cataphoric persistence, Bekins adduces evidence that DOM positively correlates with topicality. For instance, he shows that overtly coded direct objects are far more likely to be mentioned again in the subsequent discourse (i.e. they show high cataphoric persistence), since they are highly accessible and identifiable (see Arcodia and Iemmolo for similar findings on Mandarin Chinese).

The link with topicality is corroborated by the fact that *'et* is also used with temporal and spatial expressions (6.4a, b), as well as with *nominativi pendentes* (i.e. hanging topics, 6.4c) (Elwolde 1994: 175-176; Waltke and O'Connor 1990: 180-181)<sup>1</sup>. As will be shown in the remainder of this chapter, this polysemy is quite common cross-linguistically:

#### (6.4) Biblical Hebrew (Afro-Asiatic, Semitic)

- (a) *ūz<sup>ε</sup> ra'tæm 'et haššānā hašš<sup>ε</sup> mīnīt*  
 and:sow:2PL.M DOM the:year.F the:eighth.SG.F  
 "You are to sow the eighth year" (Lev. 25:22; cited in Elwolde 1994: 175)
- (b) *hēm yāsə' -ū 'et-hā-šīr*  
 3PL.M go\_out.PFV-3PL.M DOM:the:city  
 "They went out of the city" (Gen. 44:4; cited in Waltke and O'Connor 1990: 180)

<sup>1</sup>I wish to thank Eleanor Coghill for transliterating and glossing the examples (6.4a, b)

- (c) *wə-’et-huqqōt-ay lō-hālək-ū bā-hem*  
 and-DOM-decrees-my not-go.PFV.3PL in-3PL.M  
 "And as for my decrees, they did not follow them" (Ezek 20:16; cited  
 in Waltke and O’Connor 1990: 181)

Likewise, topicality is the main factor that governs the distribution of DOM in Sahidic Coptic as well. Sahidic Coptic employs two different markers for DOM, *n* before nouns and *mmo* before pronouns (Layton 2000; Engsheden 2008), as shown by examples (6.5a) and (6.5b):

(6.5) Sahidic Coptic (Afro-Asiatic, Semitic)

- (a) *a-f-joou n-nef-hmak ša n-oueeiê e-ji*  
 PST-3SG-send DOM-his.PL-servant until the.PL-farmer.PL to-take  
*n-nef-karpos*  
 DOM-his.PL-fruit  
 "He sent his servants to the farmers to take his fruit" (Engsheden  
 2008: 327)
- (b) *a-uô mmo-f kahêu*  
 PST.3PL-put DOM-him naked  
 "They stripped him" (Engsheden 2008: 329)

The distribution of DOM in Coptic has been assumed by Egyptologists to be a result of stylistic variation (Layton 2000: 131 ff.). There are some constraints on DOM, which are known amongst Egyptologists under the label of *Stern-Jernstedt rule*. According to this rule, DOM does not appear when the direct objects is not modified by a determiner and when it is governed by the predicate *wo:š* "want, love". Interestingly, DOM is compulsory whenever the direct object is governed by verbs in the imperfective aspect (i.e. present and imperfect tenses). As for tenses other than the present and the imperfect, DOM may or may not be used. The distribution of DOM in Sahidic Coptic is summarised in table (6.2), where the grey cells indicate that DOM is optional:

	Imperfective	Perfective
Pronouns	<i>n</i>	<i>n</i>
Nouns	<i>mmo</i>	<i>mmo</i>

Table 6.1: Distribution of DOM in Sahidic Coptic

Engsheden’s (2008) study is, to my knowledge, the first that recognised the alternation found in Sahidic Coptic as a case of DOM. Based on the Coptic translation of the Gospels, Engsheden shows that DOM in Sahidic Coptic is used when

the direct object is referential (i.e. definite or at least specific) and topical. Indeed, when the direct object is not referential, Coptic employs the pronominal infinitive with the direct object attached to it, as shown by example (6.6) below.

Similarly to Bekins' findings for Modern Hebrew, also in Sahidic Coptic an overtly coded direct object is likely to be mentioned again in subsequent discourse (i.e. it shows high cataphoric persistence), as shown by example (6.6). As we have discussed in Chapter 2, high cataphoric persistence is one of the main characteristics of NPs introduced in discourse through topic shifts:

(6.6) Sahidic Coptic (Afro-Asiatic, Semitic)

*a-f-joou*      *a-f-fi*      *n-t-ape*      *n-iôhannês* *hm*  
 PST-3SG-send PST-3SG-carry DOM-the-head of-John in  
*pe-šteko*    *a-u-eine*      *mmo-s*    *hijm* *p-pinaks* *a-u-taa-s*  
 the-prison PST-3PL-bring DOM-it on the-dish PST-3PL-give-it  
*n-t-šeere*      *šêm*  
 to-the-daughter little

"He sent and beheaded John in the prison. His head was brought on a platter and given to the young girl" (Engsheden 2008: 331)

In (6.6), the overtly coded direct object "head" is referred to via an overtly coded pronoun in the first occurrence (*mmo-s*), and then it is referred back to through an anaphoric pronoun suffixed to the verb, since it has been established as a highly continuous and accessible topic.

Another fact that underpins the idea the topicality governs the presence of DOM in Sahidic Coptic is the dispreference for DOM to appear when the direct object is in a negated clause, as observed by Engsheden (2008: 341). While it is clear that topicality is the main triggering factor of DOM, yet further investigation is needed to cast light on the complex DOM system found in Coptic, especially with regard to some unexpected patterns, such as the obligatory use of DOM with imperfective tenses but not with the perfective one, which is peculiar cross-linguistically.

Let us now turn to Aramaic. As we have seen in Chapter 5, DOM in Biblical Aramaic was primarily conditioned by the humanness of the direct object, although topicality had a role too. Here we will consider data from the North-Eastern Neo-Aramaic (NENA) dialects, studied by Coghill (2010, to appear), whose work focuses on the dialect spoken in Telkepe (Iraq). Unlike many NENA dialects, which have only DOI, the dialect of Telkepe has both DOM and DOI. DOM occurs solely in conjunction with DOI, as shown by examples (6.7a). DOI can instead occur alone, as in (6.7b):

## (6.7) Neo-Aramaic (Afro-Asiatic, Semitic)

- (a) *kəm-šāqəl-lə*                                  *ta*    *barāna*  
 PST-take.3SG.M-OBJ.3SG.M    DOM    ram  
 "He took the ram" (Coghill to appear)
- (b) *kəm-šāqəl-lə*                                  *barāna*  
 PST-take.3SG.M-OBJ.3SG.M    ram  
 "He took the ram"

At a first glance, the main trigger of both DOM and DOI seems to be the definiteness of the direct object. As a matter of fact, generic or non-specific direct objects, as well as direct objects which are closely integrated with the predicate, do not bring about either DOM or DOI, as shown by example (6.8)<sup>2</sup> :

## (6.8) Neo-Aramaic (Afro-Asiatic, Semitic)

- gazādə*    *g-gazd-í-wD.l*                                  *g-gazd-i-wD*                                  *zarD l...*  
 harvesters    IMPFV-HARVEST-3PL-PST    IPFV-harvest-3PL-PST    crop  
 "The harvesters harvested. They would harvest the crop..." (Coghill to appear)

Despite the previous mention, which makes a NP identifiable by the hearer, the noun "crop" does not bring about either DOI or DOM because it has generic reference and is semantically closely integrated with the predicate (Coghill to appear: 11). The close link of DOM, DOI and topicality is confirmed by the Neo-Aramaic data. Direct objects that are in narrow focus cannot take DOM, as demonstrated by example (6.9a), in which the direct object is in focus. On the contrary, the direct object in (6.9b) bears DOM because of its topical status, having been mentioned several times in the previous discourse (see Coghill to appear for further discussion):

## (6.9) Neo-Aramaic (Afro-Asiatic, Semitic)

- (a) *mán xze-lux*                                  *tā*    *lxze-li*                                  *tómD*  
 who see-PST.2SG.M    there see-PST.1SG    Thomas  
 "Who did you see there? - I saw THOMAS"
- (b) *pəš-lə*                                  *ham-āwu zəl-lə* |                                  *plāxw*  
 become.PST-3SG.M    also-he go.PST-3SG.M    working  
*gebəd-malkD* | *ta-t-qātəl-lə*                                  *ta*    *malkD*  
 chez-king    for-that-kill.3SG.M.OBJ-3SG.M    DOM    king  
 "He too went to work with the king, in order to kill the king"

<sup>2</sup>In the examples, [l] represents an intonational phrase boundary.

Therefore, in the Neo-Aramaic dialect of Telkepe, topicality is the primary trigger of both DOM and DOI, the use of both strategies being ruled out when the direct object is in focal position. A similar pattern seems to be found in other NENA dialects, as reported by Coghill, who cites data from Christian Barwar, Jewish Urmi, and Jewish Arbel dialects. In these dialects, topicality conditions the distribution of DOM and DOI. DOM and DOI can be found even with certain indefinite objects, provided that they play a prominent role in subsequent discourse<sup>3</sup> (Khan 2008, cited by Coghill). Unlike the Telkepe dialect, however, animacy in these dialects may play a role in addition to definiteness.

Many Ethiopian Semitic languages exhibit DOM: some of them, such as e.g. Amharic and Tigrinya, have DOI as well. The distribution of DOM and DOI in Ethiopian Semitic languages is mainly regulated by definiteness: definite direct objects receive overt coding and are indexed on the verb as opposed to indefinite ones, as shown by examples (6.10a, b) from Ge'ez:

(6.10) Ge'ez (Afro-Asiatic, Semitic)

- (a) *riʔyä*            *biʔise*  
 see.PST.3SG    man  
 "He saw a man"
- (b) *riʔi-o*                            *lä-biʔise*  
 see.PST.3SG-OBJ.3SG.M    DOM-man  
 "He saw the man" (Givón 1976: 164)

Both strategies are also found in Amharic (Amberber 2005, Cohen 1936, Leslau 1995, among others), where definite direct objects (including personal pronouns, proper names, NPs with the definite suffix *-w*, quantified NPs, etc.) are obligatorily overtly coded via the object marker *-[ə]n* and optionally indexed on the verb, as in (6.11a, b). However, some generics may be overtly coded, as the direct object "man" in (6.11c):

(6.11) Amharic (Afro-Asiatic, Semitic)

- (a) *Ləmma t'ərmus-u-n*            *səbbər-ə(w)*  
 Lemma    bottle-DEF-DOM    break.PFV-3SG.M-OBJ.3SG.M  
 "Lemma broke the bottle"
- (b) *Ləmma and t'ərmus səbbər-ə*  
 Lemma    one    bottle    break.PFV-3SG.M  
 "Lemma broke one bottle" (Amberber 2005: 299)

<sup>3</sup>Compare the pervasive use of DOM in Romance languages with the indefinite marker meaning "a, one", like Italian *uno/a*. See Chapter 8 for further discussion.

- (c) *igziabiher sə-i-n bə-məlk-u fət'rər-ə*  
 God man-DOM with-image-POSS.3SG.M create-PFV.3SG.M  
 "God created man in his image" (Amberber 2005: 301)

DOI -which can also be used with indirect objects in constructions similar to the dative shift found in English- becomes mandatory whenever the direct object is not overtly expressed, as in (6.12). This is easily explainable if the highly topical nature of ellipsed NPs is taken into consideration. Indeed, as I have already argued, DOI is mainly used to encode highly accessible and continuous topics (see Givón 1983: 17):

(6.12) Amharic (Afro-Asiatic, Semitic)

- (a) *məs'haf-u yət nəw?*  
 book-DEF where be.3SG.M  
 "Where is the book?"
- (b) *almaz wəssəd-əčč-\*(iw)*  
 Almaz take.PFV-3SG.F-OBJ.3SG.M  
 "Almaz took it" (Amberber 2005: 302)

Tigrinya and Tigre also show both DOM and DOI. In both cases, DOM and DOI seems to be highly sensitive to topicality, which is operationalised in terms of definiteness constraints, as we have seen for Amharic. Tigrinya direct objects can be overtly coded by the case marker *nī-*, *nä-*, and can trigger agreement if they are definite, as shown by example (6.13a) vis-à-vis (6.13b):

(6.13) Tigrinya (Afro-Asiatic, Semitic)

- (a) *lami biŋiayi ri?iy-a*  
 cow.SG.F bull.SG.M see.PFV.3SG.F  
 "A cow saw a bull" (Kifle 2007: 5)
- (b) *?it-a lami n-ät-i biŋiay*  
 DET-3SG.F cow.SG.F DOM-DET-3SG.M bull.SG.M  
*ri?iy-a-to*  
 see.PFV.3SG.F-OBJ.3SG.M  
 "The cow saw the bull" (Kifle 2007: 5)

Kifle (2007) argues that DOM and DOI in Tigrinya differ as to their triggering factors. He observes that DOM in Tigrinya hinges on definiteness, while DOI is induced by the topicality of the direct object. That topicality is the triggering factor of DOI is corroborated by the fact that, in applicative constructions, applied

objects obligatorily induce DOI (see Kifle 2007 for further discussion). It would thus seem that DOM generalised for definite direct objects, regardless of their actual topicality in discourse, whilst DOI is still closely connected to information structure.

The data reported by Kievit and Kievit (2009), however, provide a slightly different picture of the interaction of DOM and DOI in Tigrinya. While Kievit and Kievit's data support the chief role of topicality in determining DOI, yet they also show that the correlation between DOM and definiteness is not as strict as in Kifle's view. Kievit and Kievit analysed the occurrence of DOM and DOI in a contemporary Tigrinya novel, and found that definiteness alone is not the main parameter for DOM, as 53% of definite direct objects in the text are not overtly coded via DOM. Likewise, only 76% of definite direct objects show DOI. These percentages challenge the view that DOM is not sensitive to topicality. Even though Kievit and Kievit (2007) do not explicitly take topicality into consideration as a triggering factor, still it seems that topical direct objects are more likely to bring about DOM and/or DOI than non-topical ones. For instance, specific direct objects modified by the numeral "one" trigger both DOM and DOI, as in (6.14). This may be due to the fact that these direct objects are specific and thus anchored in the discourse, and conceivably play a prominent role in the ensuing discourse:

(6.14) Tigrinya (Afro-Asiatic, Semitic)

*ni-ḥadä      säbiʔayi    k-i-ḥigiz-ä-ni*  
 DOM-one.M    man.SG    INF-IMPV.3SG-help-SUBJ-OBJ.1SG  
*ḥatit-ä-yo*  
 PFV-ask-SUBJ.1SG-OBJ.3SG.M  
 "I asked a (certain) man to help me" (Kifle 2007: 7)

Likewise, Tigre mirrors the system of Tigrinya (Jake 1980), and both DOM and DOI show a rather precise association with topicality. DOM, coded by the preposition *ʔigil*, is consistently used when the direct object has been introduced into the discourse and is subsequently commented upon, just as in Tigrinya, where indefinite-specific direct objects receive overt coding because of their topicality.

A similar situation is attested amongst other Ethiopian Semitic languages. In all these languages, DOM and DOI often co-occur and are mainly determined by definiteness, which, as we have just seen for Amharic and Tigrinya, is prompted by the topic-worthiness of the direct object. In Argobba, as well as in Silṭi and Zway, the object marker is obligatorily found on definite direct objects, whilst it may optionally occur on indefinite NPs that play an important role in the ensuing discourse (see Appleyard 2004: 4):

(6.15) Argobba (Afro-Asiatic, Semitic)

*aw-u ləǧ-u-n mǎrräqe*  
 father-DEF son-his-DOM bless:PST.3SG.M.OBJ.3SG.M  
 "The father blessed the son" (Appleyard 2004: 4)

## (6.16) Zway (Afro-Asiatic, Semitic)

*hadad-ä -nom ə-rihu*  
 some-DOM their see-PST.1SG  
 "I saw some of them" (Appleyard 2004: 5)

## (6.17) Silṭi (Afro-Asiatic, Semitic)

*libaas-a-y biiša taqeebaleet*  
 clothes-OBJ-DEF from:3SG.F receive.3SG.M.3SG.F  
 "He received the clothes from her" (Appleyard 2004: 5)

The prominent role of topicality is supported, once again, by the occurrence of the object marker with temporal and spatial expressions, as well as with *as-for* NPs. These usages of the accusative are attested in Amharic as well (6.18a,b,c). In Tigrinya, the object marker is often found with topicalised direct objects, as shown by example (6.18d):

- (6.18) (a) *bet-u-n tämalləs<sup>w</sup>all*  
 house-his-DOM return.PFV.3SG.M  
 "He has returned home"
- (b) *abbat-e sämon-u-n arf<sup>v</sup>all*  
 father-my few.days-DEF-DOM pass.away.PFV.3SG.M  
 "My father passed away recently"
- (c) *yəh-ən nəgär hullu əwnät yädärräsä*  
 this-DOM thing all truth REL-arrive.PST.3SG.M  
*yə-mäsləwotall?*  
 seem.IPFV.3SG.M-to-you  
 "As for all these things, do you really think they really happened?"  
 (Appleyard 2004)
- (d) *ʔit-i hintʃ'al wäd-u, nə-t-i*  
 ART-M.SG little boy-3SG.M.GEN DOM-ART-SG.M  
*zi-Ø-siʕib ʔask'ak'i sirah*  
 REL-SUBJ.3SG.M-follow.PRS dirty work  
*ke-y-Ø-riʔi, nab mädäk'äsiʔ-u*  
 FUT-NEG-SUBJ.3SG.M.PRS-see.PRS to bedroom-3SG.M.GEN



*sägog-Ø-o*  
 send.PFV.-SUBJ.3SG.M-OBJ.3SG.M

"As for his young son, in order that he might not see the dirty work which was about to follow, he sent him to his bedroom" (Kievit and Kievit 2009: 65)

Let us now turn to Cushitic and Omotic languages. The three Cushitic languages surveyed in this chapter, Dullay, Ts'amakko, and Kemantney, have been strongly influenced by the surrounding Semitic languages (such as Amharic): thus, unlike other Cushitic languages, they do not have a marked nominative system (i.e. a system in which the nominative is overtly coded, while the accusative is left uncoded, see König 2009 and Handschuh 2010), but exhibit a DOM system (König 2009, Tosco 1994). For instance, in Dullay (East Cushitic), DOM has become restricted to definite direct objects that are left dislocated, as shown by the following example cited by Tosco (1994: 238), based on Amborn *et al.*'s (1980) grammar:

(6.19) Dullay (Afro-Asiatic, East Cushitic)

*qawhó-n mi?é hí?í*  
 man-DOM child see.PST.3SG  
 "The man, the child saw (him)" (Tosco 1994: 238)

Ts'amakko (Cushitic, Dullay) shows one of the most intriguing patterns of DOM. DOM is coded in Ts'amakko via the marker *ka/ki* and is found on pronouns, proper names and definite topical noun phrases. Interestingly, DOM does not seem to be obligatory in any of these contexts, as shown by examples (6.20a, b), where DOM optionally codes a pronoun, and (6.20c, d), where DOM is optionally found on a common noun:

(6.20) Ts'amakko (Afro-Asiatic, Cushitic)

- (a) *?usk-akk-o ?ise bog-i*  
 dirt-SG-M 3SG.F.OBJ kill-3SG.M.UNM  
 "The dirt killed her"
- (b) *?ise-ka geeʕ-i*  
 3SG.F-DOM want-1SG.UNM  
 "I want her" (Savà 2005: 141)
- (c) *kutta maaxx-e bitam-i*  
 PRON.PROX1 bead-PL.G buy-3SG.M.UNM  
 "This one buys beads" (Savà 2005: 155)

- (d) *maax-x-e xumbi ki lass-i*  
 bead-PL.G all DOM sell-3SG.M.UNM  
 "He sold all the beads" (Savà 2005: 136)

Savà (2005: 116) argues that *ka/ki* serves to topicalise direct objects as well as adverbial expressions. In this respect, Savà argues that *ka/ki* cannot be considered an object marker, as it is found on a variety of constituents, such as spatial and temporal adverbials (see Savà 2005: 121-146 for examples). Moreover, *ka/ki* is found with subordinate clauses in pre-verbal position as well (Savà 2005: 133). While Savà does not provide any semantic characterisation of the subordinate clauses where *ka/ki* occurs, a quick examination of the examples found in the grammar shows that the all the sentences to which *ka/ki* is adjoined are *since*- and conditional clauses. As I will show below, a similar distribution is attested in the Nilo-Saharan language Kanuri. Crucially, all these meanings can be directly related to the topic-marking function of the particle *ka/ki*.

Interestingly enough, there is a context in which *ka/ki* has a somewhat puzzling function that does not fit the characterisation I have proposed so far. Indeed, *ka/ki* occurs with negative verbs when in a main clause (Savà 2005: 187). The presence of what I have characterised as a topic-marker in a negative clause is quite unexpected.

In this respect, the neighbouring language Gawwada, another Dullay variety (Tosco 2010), provides some interesting data. As Tosco (2010: 344) pointed out, the primary function of the marker *kka* in Gawwada has to do with the notion of contrast, since contrast is the meaning that all of its different usages share. *kka* in isolation is an additive particle meaning "also, too". Not surprisingly, *kka* is used as a topic-shifting device: by marking contrast against the active topic, *-kka* serves to re-activate an old one (Tosco 2010: 331-333), therefore functioning as a topic promotion device<sup>4</sup>. But contrast, as argued by Molnár (2001), is an independent concept, and can therefore be applied to topics as well as to focal elements. Indeed, *-kka* has developed a meaning of unexpectedness that clearly derives from the original contrastive function of the marker (a similar path of development is attested in Meithei, see section 5.8).

Tosco (2010: 338-339) then shows that *-kka* in Gawwada is undergoing further grammaticalisation as a negative marker. This last extension, which is completely grammaticalised in Ts'amakko, appears to have arisen from the meaning of unexpectedness/contrast. This explanation, albeit still provisional, would allow one to account for the apparent anomalous grammaticalisation that the object marker has undergone in Ts'amakko and Gawwada, which I have summarised in (6.21):

<sup>4</sup>This situation shows clear parallels in Gbe languages (Fiedler 2009), and Ewe (Ameka 2010), where the additive particle fulfils a contrastive-topic marking function.

↗ contrastive topic → topic shift → **topical object marker**

(6.21) additive particle/contrast

↘ unexpectedness marker → negative

The double -and independent- grammaticalisation process undergone by *-kka* would therefore explain why this marker exhibits such an unusual variety of functions. On the one hand, the notion of contrast has been grammaticalised into a contrastive, topic-shifting marker. In turn, this topic marker, by virtue of its frequent use and association with direct objects, has been grammaticalised into an object marker for topical direct objects.

On the other hand, since contrast can be applied to focal elements as well, the very same element acquired an unexpectedness meaning, which most probably has been brought about by the notion of contrast/shift, as we have seen in Meithei. In turn, through another inferential process, a negative meaning has arisen in the unexpectedness context, as demonstrated by Tosco (2010: 339). This inference has probably paved the way for the negative-marking usage found in Ts'amakko, where the grammaticalisation process of *ka/ki* seems to be complete.

Let us now turn our attention to the interesting situation found in Haro (Omoti) and Kemantney (Cushitic). Both languages have core case-marking (i.e. subject/agent and direct object) restricted to definite NPs, i.e. NPs that carry a definiteness marker, whereas indefinite NPs never receive overt coding with respect to case marking. Haro (Woldemariam 2009) displays an interesting distribution of case morphology. Core case marking is obligatory on personal pronouns and proper names, as well as on interrogative pronouns and NPs overtly coded for definiteness (Woldemariam 2004, 2009), as shown by example (6.22a, b). Indefinite NPs are caseless, as in (6.22c).

(6.22) Haro (Afro-Asiatic, Omoti)

- (a) *ʔísí ʔés-á ʔí-wod-ín-e*  
 she-NOM he-ACC/DOM:M 3SG.F-kill-PST-AFFMT:DECL  
 "She killed him"
- (b) *ʔísí garmá-z-a ʔí-wod-ín-e*  
 she-NOM lion-DEF:M-ACC/DOM:M 3SG.F-kill-PST-AFFMT:DECL  
 "She killed the lion"
- (c) *ʔassí moló ʔé-wong-ín-e*  
 man fish 3SG.M-buy-PST-DECL:AFFMT  
 "A man bought fish" (Woldemariam 2009: 101-119)

Besides forbidding core case marking on indefinite NPs, as shown by examples above, there is also a constraint on the occurrence of the focus marker. As a matter of fact, a definite NP, be it subject/agent or direct object, cannot be overtly coded for focus, as shown by (6.23a), in which the focused direct object does not bear the definiteness marker, as opposed to (6.23b), where the simultaneous presence of the definiteness marker and the focus marker renders the sentence ungrammatical:

## (6.23) Haro (Afro-Asiatic, Omotic)

- (a) *šeebó ʔassi-kko ʔé-ʔekk-ín-e*  
 crocodile man-FOC 3SG.M-take-PST-AFFMT:DECL  
 "A crocodile took a man" (Woldemariam 2004: 221)
- (b) \**šeebó ʔassa-z-a-kko*  
 crocodile man-DEF:M-ACC/DOM:M-FOC  
*ʔé-ʔekk-ín-e*  
 3SG.M-take-PST-AFFMT:DECL  
 "A crocodile took the man" (Woldemariam 2004: 222)

Thus, Haro seems to have grammaticalised the tendency for case marking to appear with definite, topical elements. Unlike other languages I have examined, in Haro (and also in Kemantney, see below), we find differential case marking not only with definite/topical direct objects, but with definite/topical subjects/agents too (see Woldemariam 2009: 119-121 on subject focalisation).

In a similar vein, in Kemantney (Afro-Asiatic, Cushitic) core case marking is restricted to definite NPs (Leyew 2003: 237-240) as in (6.24a) vs. (6.24b), although it does not seem clear how definiteness is coded in Kemantney, as pointed out by König (2009: 70-72):

## (6.24) Kemantney (Afro-Asiatic, Cushitic)

- (a) *k<sup>w</sup>ant-i nīṅ-is xätīy-ay*  
 ice-NOM house-DOM puncture-PST  
 "The ice pierced the house"
- (b) *an šäb ʃax-iy<sup>w</sup>*  
 1SG milk drink-PST  
 "I drank milk" (Leyew 2003: 239-241)

Since Kemantney is a moribund language, it will be nearly impossible to delve deeper into its distribution of case marking. However, it seems that, in Kemantney

too, the tendency for core arguments to be topical has been conventionalised in that only definite, topical, arguments show case-marking.

Unlike Kemantney and Haro, core case marking in Sheko (main word order SOV; Hellenthal 2010) is not restricted to definite NPs, as the subject/agent is not overtly coded. DOM seems to be influenced by definiteness, even if overtly coded indefinite direct objects are found (Hellenthal 2010: 258), as shown by (6.25a, b). As Hellenthal (2010: 259) observes, definiteness *per se* is not a necessary and sufficient condition for DOM to show up, as it is possible to have indefinite direct objects overtly coded and vice versa. Yet, there is one context in which DOM is nearly mandatory, i.e. when the direct object is in initial position, as in (6.25c). Thus, in Sheko too, the correlation between dislocation and DOM holds again.

(6.25) Sheko (Afro-Asiatic, Omotic)

- (a) *dēyḡn̄*      *yí-fyááyń-əra*                      *kòb-tə*  
 girl.F.DEF 3SG.F.POSS-frog.F.DEF-DOM take-SS  
 "The girl took the frog and..."
- (b) *kútfî-ra*              *há-buuts'u-k-ə*  
 chicken-DOM 3SG.M-pluck-REAL-STI  
 "He plucked a chicken"
- (b) *há-şóóş-ń-ş-ara*                      *yááb-m-s*      *dufu-t-á*              *wuş-ń*  
 3SG.M-snake-DEF-M-DOM man-DEF-M hit-SS-3SG.M kill-DS  
 "The man hit the snake and killed it" (Hellenthal 2010: 259)

## 6.3 Nilo-Saharan

As observed by Dimmendaal (2010), despite the fact that DOM is widespread among the languages of Africa, it has been ignored in the typological literature. Indeed, apart from the notable exceptions of Bantu and Afro-Asiatic languages, which have been the subject of an increasing amount of studies (see sections 6.4 and 6.2 respectively), the growing amount of data drawn from Nilo-Saharan and other Niger-Congo languages have virtually gone unnoticed outside the domain of Africanists. In this section, then, I will attempt to provide a general picture of the DOM and DOI systems in Nilo-Saharan languages.

The Nilo-Saharan languages that will be examined in this chapter are: Fur, Maba, Tama, Kanuri, Tubu, Dongolese Nubian, and Uncunwee. Most of the data for these languages come from grammatical descriptions and the excellent survey presented by Dimmendaal (2010).

An interesting case of DOM comes from Fur (Jakobi 1989, Waag 2009 cited in Dimmendaal 2010). Fur is a verb-final language (with both SOV and OSV word

orders found) with no case marking on core arguments besides direct objects. Pronominal direct object are obligatorily overtly coded (see examples in Jakobi 1999: 93, and Dimmendaal 2010). At first glance, DOM, expressed by the clitic *ʃi* used also for recipients and beneficiaries, seems to be conditioned by definiteness and specificity, as indefinite non-specific direct objects never bring about DOM, as in (6.26a) vs. (6.26b):

(6.26) Fur (Nilo-Saharan, Fur)

(a) *ká bá wʒel ʔáɫáŋ bəwtê-iŋ para dig*  
 1SG just 1SG.want.IPFV to hoe.GEN handle one  
*ʔandi*  
 1SG.cut.SUBJ  
 "I just want to cut a handle for a hoe"

(b) *namá k-in pʊttɔŋ-ʃi tɔŋ ʔ-ʊndí*  
 then PL-this dry.PL-DOM house.LOC 1SG-leave.PFV  
 "Then I should leave the dry ones at home" (Waag 2009, cited in Dimmendaal 2010: 19)

Interestingly, DOM is more likely to be used when direct objects are topicalised (i.e. when they do not occur immediately pre-verbally), as shown by examples (6.27). In (6.27a), the direct object is topicalised: on the contrary, as Dimmendaal points out, when "no such contrastive function is involved", the direct object is left uncoded, as in (6.27b):

(6.27) Fur (Nilo-Saharan, Fur)

(a) *nârma-ʃi namá lia k-âm kí-dulê's tɔkke*  
 food\_in\_market-DOM then later 1PL-eat.SUBJ with-sun hot  
 "Eating at the market we can do later on, when it is midday"

(b) *nuŋ b-ámu-lá*  
 food 2PL-eat.COMPL\_Q  
 "Have you eaten?" (Waag 2009, cited in Dimmendaal 2010: 20)

Most probably, the contrastive function to which Dimmendaal refers has to be intended as a topic promotion device used to signal a contrast/shift against a set of previous referents (see Prince 1997). An interesting example is (6.28), where we have multiple instances of NPs coded with accusative. While one of these NPs is the addressee of the predicate "say", the other two NPs seem to be topics, as also shown by the translation. As we have already seen, that the object marker has an "aboutness" meaning seems to be quite widespread cross-linguistically:



## (6.31) Uncunwee (Nilo-Saharan, Eastern Sudanic)

- (a) *nò ìdō tòwààŋ-\*(gì) kón-òŋ*  
 this woman child-DOM have.INALIEN-3SG.PRS  
 "The woman has a child (i.e. She is pregnant)"
- (b) *ìdó tòndò-ò kwāŋ-óŋ*  
 woman child-DOM have.ALIEN-3SG.PRS  
 "The woman has the (born) child" (Jakobi 2009: 8)

In ditransitive clauses, the object marker may encode both the direct and the indirect object, as shown by (6.32)<sup>5</sup>. The most common situation, however, seems to be one in which both are left uncoded (with the fixed order indirect object/direct object).

## (6.32) Uncunwee (Nilo-Saharan, Eastern Sudanic)

- yě tòndó-‘ò kāmé-é ĩ-ēbé*  
 1SG.SUBJ child-DOM food-DOM give-1SG.PST  
 "I gave food to the child" (Jakobi 2009)

The other three Eastern Sudanic languages surveyed here, Tama, Maba, and Nubian, have not been the subject of in-depth enquiry with regard to the organisation of their case-marking systems. The data available on Tama (Dimmendaal 2009, 2010) shows a rather intricate DOM system. The object marker *-iŋ* (used also for indirect objects and beneficiaries) seems to be obligatory with pronouns and proper names, as in (6.33a), while it is optional with other direct objects, regardless of whether they are definite or not<sup>6</sup>. Dimmendaal (2010) states that human definite direct objects are more likely to induce DOM, as in (6.33b). Nonetheless, DOM is found with some inanimates as well, as in (6.33c, d). In the latter example, the object marker is optional:

## (6.33) Tama (Nilo-Saharan, Eastern Sudanic)

- (a) *wâ-ŋ áwí tíí<sup>4</sup>ní-ŋá*  
 1SG-DOM snake bite.3SG.PFV  
 "A snake bit me"
- (b) *híná í-r tààt-ír-<sup>4</sup>íŋ lóó<sup>4</sup>wéy*  
 my wife child-DEF-DOM 3SG.drink.CAUS  
 "My wife is feeding the baby"

<sup>5</sup>It is also possible that either the direct or the indirect object is overtly coded (Jakobi 2009: 5).

<sup>6</sup>Definiteness may be coded by the marker *-iŋ*, as in (6.33b).



- (c) *dòktîr léék-<sup>1</sup>îŋ fàsúón<sup>1</sup>éy*  
 doctor urine-DOM check.3SG  
 "The doctor will check the/your urine"
- (d) *èsì àna<sup>1</sup>ár-(îŋ) fú<sup>1</sup>té*  
 3SG ground.DEF-DOM sweep.3SG  
 "(S)he is sweeping the floor" (Dimmendaal 2009: 320-325)

According to Dimmendaal (2010: 24-25, who quotes Satti 2008), the distribution of DOM in Dongolese Nubian seems to be quite similar to Tama. In the examples he provides, DOM is found with a definite direct object (6.34a); interestingly, the marker appears both on the direct and the indirect object when both are animate (6.34b):

(6.34) Dongolese Nubian (Nilo-Saharan, Eastern Sudanic)

- (a) *buru kusu-gi kandi kinynya-ged mer-ko*  
 girl meat-DOM knife small-INSTR cut-PST  
 "The girl cut the meat with the small knife"
- (b) *ay tii-gi essi-gi tirsi*  
 I cow-DOM water-DOM give.1SG.PST  
 "I gave water to the cow" (Dimmendaal 2010: 24-25)

As argued by Dimmendaal (2009: 325, 2010), besides animacy/definiteness, information structure distinctions seem to play a role in the appearance of DOM in Tama and Nubian. However, the available data does not make it possible to determine which factor(s) take(s) priority over the others in determining the distribution of DOM.

Maba has quite a complex DOM system, which interacts with a DOI system. First of all, it appears that a dedicated object marker is found only in the plural (Weiss 2009: 168). Indeed, the definite subject/agent and the direct object are overtly coded in the same way by means of the clitic *-gu* for the singular and *-nu* for the plural, as in (6.35a). The definite plural direct object is overtly coded by means of the clitic *-nú-gù*, as in (6.35b). The situation, however, is made more complex by the fact that independent object pronouns are obligatorily overtly coded by the clitic *-gú/gù*, as in (6.35c, d), and indexed on the verb as long as they refer to the 1/2 person (both singular and plural, see Weiss 2009: 226):

(6.35) Maba (Nilo-Saharan, Eastern Sudanic)

- (a) *ŋû:-gù sàgàr-gù t-ìŋgèr-í-r-í*  
 dog-SG.DEF jackal-SG.DEF 3SG-ask-PST-PL-DECL  
 "The dog asked the jackal" (Weiss 2009: 169)

- (b) *àmárà-gù kól-í:nú-gù t-íggè:rì*  
 lion-SG.DEF child-PL.DEF-DOM 3SG-ask.DECL  
 "The lion asks the children..." (Weiss 2009: 170)
- (c) *àm-gú ànd-ùr-í-r-ì*  
 1SG-DOM 1SG.OBJ.2/3SG.SUBJ-call-PST-PL-DECL  
 "(As for me), he called me"
- (d) *káj-gú kùnd-ùlsí-tè*  
 2PL-DOM 2PL.OBJ.3SG.SUBJ-wait-FUT  
 "(As for you), he will wait for you" (Weiss 2009: 172)

Maba has an SOV basic word order, and initial position serves to topicalise a constituent (Weiss 2009: 293, 396). Therefore, it seems that, once again, a relationship between dislocation and presence of DOM does exist.

The link between dislocation and presence of DOM is attested in Tubu and Kanuri as well. DOM in Tubu (basic word order SOV) is obligatory with personal pronouns, as shown by example (6.36a)<sup>7</sup>. Although Lukas states that DOM is governed by definiteness, apparently, the marker may be absent even though the direct object is highly definite (see discussion in König 2009: 40 ff.). However, there is a preference for DOM to appear when the direct object is dislocated in initial position, as in (6.36c):

(6.36) Tubu (Nilo-Saharan, Saharan)

- (a) *səŋgá ga gòyintu*  
 3SG DOM take.3PL.A  
 "They take him"
- (b) *dugulí maná ga wáu*  
 lion squirrel DOM hit.3SG.A  
 "The lion hit the squirrel"
- (c) *dugulí ga ài-kεε du maná yi tárwayi*  
 lion DOM so POST squirrel NOM outwit.3SG.A  
 "So the squirrel outwits the lion" (Lukas 1953: 161, cited in König 2009: 40)

Interestingly, the object marker is reported to have also a topic marking function, as shown by example (6.37). Furthermore, as König (2009) discusses, Lukas (1953: 161) paraphrases its function as "as for", i.e. as having an "aboutness" meaning. Even though König finds it highly unusual that a topic marker develops into an object marker, this polysemy seems to be quite common cross-linguistically:

<sup>7</sup>Glosses and translations of the original examples are taken from König (2009).

## (6.37) Tubu (Nilo-Saharan, Saharan)

*gadú nta ga*  
 warthog you DOM  
 "Oh warthog! As for you..." (Lukas 1953: 161)

Kanuri (basic word order SOV) shows an even more conspicuous link with topicality. DOM appears to be obligatory with independent object pronouns, otherwise it is optional. Overt (both subject and object) pronouns are used only when they need special emphasis, as in dislocations, otherwise pronominal arguments (only 1/2 person) are signalled on the verb. It is however possible to have both strategies, as in (6.38a), where the (dislocated) independent object pronoun receives DOM and is indexed on the verb (Hutchinson 1981: 212).

Hutchinson (1981: 211 ff.) further argues that the use of DOM with other kinds of NPs is never used with non-animate direct objects. With animate direct objects, DOM is optional, but more likely when direct objects are dislocated in initial position and thus constitute the (contrastive) topic of the sentence, as in (6.38b, c):

## (6.38) Kanuri (Nilo-Saharan, Saharan)

- (a) *Nyí-à nzúrukə-ná*  
 2SG-DOM see.PST.1SG-2SG  
 "You, I saw you" (Hutchinson 1981: 212)
- (b) *tádànzə-á cúrò*  
 child-DOM see.PST  
 "Her/his child, I saw him" (Hutchinson 1981: 215)
- (c) *Ali-a suruna*  
 Ali-DOM see.PST  
 "He/she saw Ali" (Hutchinson 1986: 195)

It is noteworthy that the Kanuri object marker shows other uses that are strongly connected with topicality. Indeed, Hutchinson (1981: 247) explicitly states "any NP may be marked as a topic through the application of the associative postposition (i.e. the object/topic marker GI) to set it off in construction-initial position" (Hutchinson 1981: 247). Here are some examples:

## (6.39) Kanuri (Nilo-Saharan, Saharan)

- (a) *wú-à nòngənuí*  
 1SG-DOM know.NEG.1SG.PRS  
 "As for me, I don't know" (Hutchinson 1980: 343)

- (b) *kû-à*            *lènginbâ*  
 today-DOM go.1SG.NEG.IPFV  
 "As for today, I won't go" (Hutchinson 1980: 343)
- (c) *shí mâi-à wú yé gâlìwùngin*  
 3SG king-DOM 1SG too get\_rich.1SG.IPFV  
 "If he is the king, I too will become rich" (Hutchinson 1980: 346)

These examples are in accordance with the cross-linguistic data I have analysed thus far, with the object marker having or having had a topic marking function. Kanuri displays other interesting facts, which deserve further investigation. For example, as shown by (6.39c), the same postposition used as object marker is deployed to encode verbless conditional and causal-*since* clauses (Hutchinson 1981: 290-292: see also Cyffer 1983).

As is well known, conditionals share many features with topics: they often have the same distribution at sentence initial position and the same information status, being both presupposed (Haiman 1978: 567). In a similar manner, the content of *since*-clauses is usually presupposed, as argued by Lambrecht (1994: 69-70). Therefore, it may well be that the object marker and the conditional/causal marker derive from the same historical source. Of course, further data are needed in order to ascertain whether this correlation does hold in other languages and the possible grammaticalisation paths that the postposition might have followed.

## 6.4 Niger-Congo

In this section I will analyse data from Niger-Congo languages. Bantu languages in particular have been the subject of many studies, which will be nearly impossible to review here. As is well known, Bantu languages display DOI, which has been taken as being determined by topicality and animacy (see Duranti 1979, Hyman and Duranti 1979, Bresnan and Mchombo 1987, Wald 1979, to name only a few).

However, the amount of variation found among Bantu languages is very large, and in many cases, as we will see shortly, the direct link with topicality has been lost in the course of the grammaticalisation of the construction. I will restrict my attention to monotransitive clauses, as ditransitive clauses display some additional issues that cannot be discussed here, such as the order of affixes on the verb, the number of object markers allowed per sentence, etc. (see Bresnan and Moshi 1990, Beaudoin-Lietz et al. 2004 for further information).

The distribution of DOI in Chichewa is directly related to topicality, as shown by Bresnan and Mchombo (1987). Chichewa shows both subject and object indexation in its verbal morphology. In finite verb forms subject/agent indexation

is obligatory, while DOI is optional. While the subject/agent index is ambiguously used for grammatical and anaphoric agreement, DOI is invariably used as a topic-anaphoric device and, as we will see shortly, is analysed as an incorporated pronominal argument by Bresnan and Mchombo (1987: 745). They suggest that this is so based on word order evidence. As a matter of fact, when there is no DOI, the direct object must obligatorily follow the verb and cannot be freely moved, as shown by (6.40a) as opposed to (6.40b). By contrast, when there is DOI, all word order permutations are possible:

(6.40) Chichewa (Niger-Congo, Benue-Congo)

- (a) *Njûchi zi-ná-lúm-a alenje*  
 bee.PL SUBJ-PST-bite-IND hunter.PL  
 "The bees bit the hunters"
- (b) \**Alenje zi-ná-lúm-a njûchi*  
 hunter.PL SUBJ-PST-bite-IND bee.PL  
 "The bees bit the hunters"
- (c) *Njûchi zi-ná-wá-lúm-a alenje*  
 bee.PL SUBJ-PST-OBJ-bite-IND hunter.PL  
 "The bees bit them, the hunters"
- (d) *Alenje njûchi zi-ná-wá-lúm-a*  
 hunter.PL bee.PL SUBJ-PST-OBJ-bite-IND  
 "The bees bit them, the hunters" (Bresnan and Mchombo 1987:  
 744-745)

Bresnan and Mchombo correctly observe that DOI is prohibited to co-occur with direct objects that are not topical. Indeed, DOI is systematically ruled out when the direct object is a *wh*-phrase, a cognate object or part of an idiom or light verb construction, and in general with focal direct objects (Bresnan and Mchombo 1987: 759-764). From a syntactic point of view, they assume that there is a difference in the syntactic status of the direct object NP. As a matter of fact, they argue that, in the absence of the OM, the direct object NP must remain in the post-verbal position, being thus inside the VP (Bresnan and Mchombo 1987: 745). On the other hand, when there is DOI, the NPs display freedom of position, and the NP is analysed as a floating topic outside the clause boundary. Thus, in their view, the object indexation marker is an incorporated pronominal argument that satisfies the argument structure of the verb.

Bresnan and Mchombo's claim that direct objects triggering DOI in Chichewa are VP-external is consistent with the frequent connection between dislocation and DOM we have found in many typologically diverse languages (for a detailed

critique of this hypothesis, see Riedel 2009). Indeed, in at least another Bantu language, Kinande (Baker 2003), DOI is consistently used only when the direct object is left-dislocated, as shown by the ungrammaticality of (6.41a) vis-à-vis (6.41b):

(6.41) Kinande (Niger-Congo, Benue-Congo)

- (a) *N-a-(\*)ri-gul-a eritunda*  
 1SG.SUBJ-TM-OBJ-buy-FV fruit  
 "I bought a fruit"
- (b) *Eritunda n-a-(\*)ri-gul-a*  
 fruit 1SG.SUBJ-TM-OBJ-buy-FV  
 "The fruit, I bought it"

Similar patterns are attested in Setawana (Demuth and Johnson 1989), Setswana (Demuth 1989), and Zulu (Buell 2006). In these languages, DOI is found only with topical direct objects, i.e. when they are dislocated. In addition, in Setswana and Setawana, DOI is prohibited with inherently focal elements, such as question words. Dzamba (cited by Bresnan and Mchombo 1987, see Bokamba 1981, 1992) shows an interesting distribution of the object marker when the direct object is in initial position. A preposed direct object (i.e. topicalised) apparently does not bring about DOI, while a left-dislocated direct object does.

However, Bresnan and Mchombo's view turns out to make incorrect predictions as to the complementarity of case marking and indexation. They predict that if a language has both object marking and object indexation, the topical direct object indexed by the topic-anaphoric pronoun should not be subject to case assignment as it is outside the VP and is linked to the verb via an anaphoric relation (Bresnan and Mchombo 1987: 765). Thus, case and indexation should be in complementary distribution. But the data I have examined so far consistently shows that dislocated direct objects often receive DOM, and in some languages DOM seems to be restricted to such direct objects. In all these cases, there is no evidence to assume that case assignment to these direct objects is not governed by the verb, as we have argued in Chapter 4 with regard to the emergence of DOM on dislocated direct objects.

Thus, all the Bantu languages present strong evidence for the theory that DOI (and in general indexation systems) arise from topic-shift constructions (like dislocations) in which there is a pronoun that refers anaphorically to the dislocated NP, as proposed by Givón (1976, 1979) in his seminal work. According to his view, the anaphoric pronoun undergoes phonetic and morphological reduction as a result of overuse, and then the whole construction comes to be reanalysed as

pragmatically neutral, with the reduced pronoun/clitic used as a true indexation marker.

Bantu languages are at different stages in the grammaticalisation of DOI. While DOI is still mainly governed by the topicality of the direct object in the languages I have discussed above, in other languages DOI has lost its direct connection with topicality and has been extended to direct objects with topic-worthiness features. In Swahili, for example, DOI is nearly obligatory with animate direct objects (including the question word "who"), while it is optional with inanimate direct objects (Wald 1979, 1997; Seidl and Dimitriadis 1997; Marten and Kula ms). Nonetheless, when DOI is used with an inanimate direct object, there is a difference in definiteness, as the direct object must be interpreted as definite:

(6.42) Swahili (Niger-Congo, Benue-Congo)

- (a) *ni-li-mw-on-a*  
 SUBJ.1SG-PST-OBJ-see-FV  
 "I saw him/her"
- (b) *ni-li-\*(mw)-on-a*                      *Juma*  
 SUBJ.1SG-PST-OBJ-see-FV    *Juma*  
 "I saw Juma"  
*ni-li-on-a ki-tabu*
- (c) *ni-li-(ki)-on-a*                      *kitabu*  
 SUBJ.1SG-PST-OBJ7-see-FV    *book*  
 "I saw (a)the book" (Marten and Kula 2007: 5)

Similarly, in Sambia, DOI is required with first/second person pronouns, proper names, and some kinship terms and optional with other human/animate direct objects; sometimes, inanimate direct objects can trigger DOI too (Riedel 2009: 44 ff.). As in Swahili, DOI is required with the *wh*- word "who". The occurrence of DOI with the *wh*-word is easily explainable in terms of extension to direct objects which have topic-worthiness features, such as animacy, regardless of the actual topicality. The cases of Makhwa and Ngiemboon, in which DOI is limited to some noun classes (usually human/animate) and pronouns (Stucky 1981, van der Wal 2009: 84 ff. for Makwua, Watters 2003: 249 for Ngiemboon) regardless of the actual animacy/definiteness/topicality of the direct object, can be accounted for in a similar manner.

This, however, does not imply that the link with topicality is completely lost in synchrony. Indeed, clear topicality effects on DOI are still easily observable in languages, like Swahili, where DOI has been losing its direct connection with topicality. For instance, as noted by Wald (1979, 1997), pre-verbal (i.e. topical)

direct objects are still more likely to bring about DOI than post-verbal ones in contemporary Swahili<sup>8</sup>. Conversely, as I have argued above, DOI in other Bantu languages is still more closely connected to the topical position of the direct object.

I conclude this section on Niger-Congo languages by examining the distribution of DOM in Dogon (Plungian 1995). DOM is obligatorily employed for pronouns, proper names, and kinship terms (6.43a, b, c):

(6.43) Dogon (Niger-Congo, Dogon)

- (a) *Sana kanda \*(ɲ) bo-e-∅*  
 Sana Kanda DOM call-AOR-3SG  
 "Sana called Kanda"
- (b) *Sana mi \*(ɲ) bo-e-∅*  
 Sana 1SG DOM call-AOR-3SG  
 "Sana called me"
- (c) *Sana nna \*(ɲ) bo-e-∅*  
 Sana mother DOM call-AOR-3SG  
 "Sana called his mother" (Plungian 1995: 12)

The clitic *ɲ* is also used to code the indirect object. In addition, it can be employed to signal contrast, as shown by (6.44):

(6.44) Dogon (Niger-Congo, Dogon)

- (a) *gamma gɛ ay aw-e-∅*  
 cat DEF mouse catch-AOR-3SG  
 "The cat caught a mouse"
- (b) *gamma gɛ ay nɔ ɲ aw-e-∅*  
 cat DEF mouse this CNTR catch-AOR-3SG  
 "This mouse, the cat caught it" (Plungian 1995: 13)

Once again, the object marker shows a close link with an information-structure category, contrast, which can be one of the sources of object markers. However, more data is needed to clarify the connection between the distribution of the marker and information structure. As we have seen, the situation found in Dogon somewhat mirrors the situation found in Ts'amakko, where the object marker is also used to signal contrast and unexpectedness.

<sup>8</sup>See also Seidl and Dimitriadis (1997) for a corpus study of DOI in Swahili based on Prince's (1992) definitions of information status. Their results corroborate the view that topicality does still play a role in the distribution of DOI in Swahili.



## 6.5 Indo-European and Dravidian

In this section I will analyse the DOM systems found in (Modern Eastern) Armenian, Persian and Balochi. Let us start with Armenian. There are two standardised written varieties of Modern Armenian, Eastern and Western Armenian, which have developed since the 18th century (Dum-Tragut 2009: 1)<sup>9</sup>. Both of them have DOM, which uses the same marker employed to indicate the dative. Dum-Tragut (2009: 374) provides the following characterisation of DOM in Modern Eastern Armenian:

1. (+human), (+animate), definite (as in (6.45a))
2. (+human), (+animate), indefinite (as in (6.45b))
3. personificated (-human/-animate) nouns regardless of whether they are definite or indefinite (as in (6.45c))

(6.45) Eastern Armenian (Indo-European, Armenian)

(a) *T'urk'-er-ě irenc' amena-t'ank ban-ě*  
 Turk-PL.NOM-the their most\_expensive thing.NOM-the  
*erexa-ner-i-n vstah-um en*  
 child-PL-DOM-the entrust-PTCP.PRS they.are  
*hayastanc'i-ner-i-n*  
 Armenian-PL-DOM-the  
 "The Turks entrust their most precious thing, the(ir) children, to Armenians"

(b) *menk' jeřnamux enk' el-el šat "eřand*  
 we.NOM disposed are be-PTCP.PFV many energy.NOM  
*c'uc'aber-ac" mardk-anc' gtn-el*  
 show-PTCP.RES person-PL.DOM find-INF  
 "We have disposed ourselves to find many persons showing/having shown energy"

(c) *Ani-n sir-um ē ir hor*  
 Ani.NOM-the love-PTCP.PRS be.3SG.PRS.F her father.DAT  
*šan-ě*  
 dog-the  
 "Ani loves her father's dog" (Dum-Tragut 2009: 374)

<sup>9</sup>We will discuss here only Eastern Armenian. See Nilsenova (2002) for a discussion of DOM in Western Armenian

However, Megerdooonian (1999) provides many examples that do not fit in this characterisation, as in many cases the overtly coded direct objects are inanimate but definite, as in (6.46a). Interestingly, Megerdooonian shows that there is a striking difference in stress patterns depending on the presence of overt coding. When a direct object is overtly coded, the main sentence stress is assigned to the verb only, and the direct object is definite, as in (6.46b). Uncoded direct objects instead systematically attract stress, as shown by (6.46c):

(6.46) Eastern Armenian (Indo-European, Armenian)

- (a) *Ara-n vst'ah ays girk-ə k'-k'arta*  
 Ara-NOM certainly this book-DOM PTCP-read.3SG  
 "Ara will certainly read this book"
- (b) *Ara-n girk-ə AYR-ETS*  
 Ara-NOM book-DOM burn-3SG.AOR  
 "Ara BURNT the book"
- (c) *Ara-n GIRK gn-ets*  
 Ara-NOM book buy-3SG.AOR  
 "Ara bought A BOOK/BOOKS" (Megerdooonian 1999: 313-314)

This pattern can be interpreted as a reflection of the topicality of the direct object. When the direct object is the argument focus, it does not induce DOM, and is therefore accented. To the contrary, when the direct object is part of a predicate-focus structure, it is the verb that is accented, whereas the overtly coded direct object is de-accented because of its presupposed (thus topical) nature. Indeed, as I have argued in Chapter 4, it is possible for topical elements to be found within the focal domain.

Classical Armenian had DOM too, though this kind of DOM has been lost in both varieties of Modern Armenian. DOM in Classical Armenian was primarily determined by the definiteness of the direct object (Meillet 1903: 93 ff., Thomson 1989) and, unlike Modern Armenian, was coded by the preposition *z-*. As noted by Meillet (1903: 93) and Bubenik (2006: 161), the preposition employed to overtly code definite direct objects had originally the meanings "concerning, (a)round", resembling thus a topic marker. It is thus likely that DOM in Classical Armenian was conditioned by topicality as well, even though no studies have dealt with this specific aspect of the distribution of *z-*.

Let us turn to Persian, where the development and the synchronic distribution of DOM have been the topic of several studies. DOM in Persian is commonly said to be determined by definiteness, as shown by the following minimal pair:

(6.47) Persian (Indo-European, Indo-Iranian)

- (a) *Hasan ketab-râ did*  
 Hasan book-DOM see:PST.3SG  
 "Hasan saw the book"
- (b) *Hasan ketab did*  
 Hasan book see:PST.3SG  
 "Hasan saw a book" (Comrie 1989: 132)

However, the usage of *-râ* and its variants *-o*, *-ro* is more complex. First of all, *-râ* is obligatory with pronouns, proper names and direct objects modified by demonstratives. These uses are easily predictable, since these NP categories are definite *par excellence*. However, DOM can co-occur with direct objects that bear the indefiniteness marker *-i-*, as in 6.48:

(6.48) Persian (Indo-European, Indo-Iranian)

- xane-i-ra atiš-zæd-ænd*  
 house-IND-DOM burn-struck-3PL  
 "They burnt a certain house" (Mahootian 1997: 198)

Many scholars (Karimi 1990, Lazard 2001, among others) have invoked the notion of specificity (see Chapter 4) in order to account for examples like (6.48). Therefore, the cut-off point of DOM in Persian would include indefinite direct objects, as long as they are specific. Indeed, indefinite non-specific direct objects do not induce DOM, as in (6.49). As noted by Karimi (1990: 148), specificity can be easily identified by means of anaphoric pronouns: if the direct object is specific (and overtly coded), the anaphora must be coded by the pronoun "it" (6.49c), otherwise the indefinite pronoun "one" will be used (6.49b):

(6.49) Persian (Indo-European, Indo-Iranian)

- (a) *diruz ru miz ye sib gozâšt-am*  
 yesterday on table a apple put.PST-1SG  
 "I put an apple on the table yesterday" (Karimi 1990: 140)
- (b) *Râmin pirhan xarid man ham yeki xarid-am*  
 Ramin shirt buy.PST I also one buy.PST-1SG  
 "Ramin bought a shirt, I bought one too"
- (c) *man mi-xâst-am pirhan-e sabz-o be-xar-am*  
 I PRS-wanted-1SG shirt-EZ green-DOM SUBJV-buy-1SG  
*ammâ Râmin un-o zud-tar xarid*  
 but Ramin it-DOM soon-COMP buy.PST  
 "I wanted to buy the green dress, but Ramin bought it first" (Karimi 1990: 148)

Thus, only indefinite specific direct objects can be overtly coded, while indefinite non-specific ones cannot receive DOM. Recall from Chapter 4 that specificity serves to provide a referential anchoring of the indefinite specific NP to another entity expression (i.e. the anchor) in discourse (von Heusinger to appear). Thus, indefinite specific direct objects can be construed as topical, since they are identifiable and informationally salient. That DOM in Persian has a close link with topicality is an idea that has been repeatedly advanced in the last three decades (Peterson 1974, Windfuhr 1979, 1990; Dabir-Moghaddam 1992, Ghome-shi 1997). There is, indeed, both synchronic and diachronic evidence that DOM is closely connected with topicality. In what follows, I will present the synchronic data supporting the topicality hypothesis.

Similarly to other languages I have been discussing in this chapter, Persian *-râ* may be found on left dislocated constituents, such as direct objects (6.50a), indirect objects (6.50b), or obliques (6.50c)<sup>10</sup> :

(6.50) Persian (Indo-European, Indo-Iranian)

- (a) *unja-ro ne-mi-xa-m to be-bin-i-š*  
 there-DOM NEG-DUR-want-1SG you SBJV-see-2SG-3SG.PC  
 "That place, I don't want you to see it" (Mahootian 1997: 124)
- (b) *irj-o pul be-heš be-d-e*  
 Iraj-DOM money to-3SG.PC IMP-give-3SG  
 "Iraj, give him money" (Mahootian 1997: 124)
- (c) *man-o beh-me mi-xand-e*  
 I-DOM at-1SG IMPF-laugh-3SG  
 "She laughs at me" (Karimi 1990: 144)

In addition, *-râ* is frequently found on spatial and temporal expressions (6.51a, b), as well as on possessors or elements that are in a part-whole relationship (6.51c), as shown by the examples below:

(6.51) Persian (Indo-European, Indo-Iranian)

- (a) *emšab-o kamāl injā mi-mān-e*  
 tonight-DOM Kamal here IPVF-remain.PRES-3SG  
 "Tonight, Kamal is staying here" (Mahootian 1997: 121)
- (b) *ta xune-ro dovid-m*  
 until house-DOM run.PST-1SG  
 "I ran till home" (Karimi 1996)

<sup>10</sup>Persian basic word order is SOV.

- (c) *mašin-o dar-eš-o bast-am*  
 car-DOM door-POSS.3SG-DOM close.PST-1SG  
 "As for the car, I closed its door" (Karimi 1990: 143)

As we have seen in other languages, all the constituents overtly coded by *-râ* share a high degree of topicality. This high topicality is indicated by the structural position they occupy, as left dislocations in Persian are a means of topicalisation (Mahootian 1997: 124), and by the presence of the postposition<sup>11</sup>.

The important role played by topicality is, once again, upheld by diachronic evidence. As shown by Windfuhr (1987, 2009), *-râ* derives from the Old Persian postposition *râdi* "by reason of, concerning". This postposition became *rây* in Middle Persian and was used to express cause and reference, with a meaning similar to "as for" in English. In Late Middle Persian, it then started being employed with topicalised indirect and direct objects.

DOM in Contemporary Persian, however, is not governed by topicality alone. As a matter of fact, DOM has grammaticalised as a marker of direct objects which show topic-worthiness features, such as definiteness/specificity and animacy. This explains why pronouns, proper names and inherently definite NPs must be obligatorily overtly coded, irrespective of their topical status. Since it is partly based on topic-worthiness features, DOM has been extended to prototypical focal elements as well, such as the question words *ki* "who", *ci* "what", and *kodum* "which", as in (6.52):

(6.52) Persian (Indo-European, Indo-Iranian)

- (a) *Ki-ro emruz did-i*  
 who-DOM today see.PST-2SG  
 "Who did you see today?"
- (b) *kodum ketâb-o xund-i*  
 which book-DOM read.PST-2SG  
 "Which book did you read?" (Karimi 1996: 154)

I will discuss more thoroughly this process of extension in Chapter 8, where the development of DOM in Romance languages will be examined.

A DOM system similar to that found in Persian is attested in Persian Tajik (Windfuhr and Perry 2009: 485-486) and in Wakhan Wakhi (Bashir 2009: 844), where definite and specific direct objects are overtly coded by a cognate of the Persian direct object marker (*-râ/ro*). In other Iranian languages, such as Balochi,

<sup>11</sup>As noted by Karimi (1990: 152), the topical status of these NPs is further corroborated by the fact that they do not bear stress, they are subject to resumptive pronominalisation (except for adverbials), and they denote old (i.e. presupposed) information.

Vafsi and Hunza Wakhi, we find a much more complicated system, since DOM interacts with the ergative systems present in these languages. Here I will discuss some data from Balochi, which are primarily gleaned from Korn (2008). Balochi has a split-ergative system based on tense, with present verb forms following a nominative-accusative system and past verb forms following an ergative one (Korn 2008). Balochi has a case system that contrasts a direct case (morphologically zero) with an oblique case (sing. *-ā*, pl. *-ān*) and an accusative case (sing. *-ārā*, pl. *-ānā -ānrā*). Indefinite direct objects occur in the direct case (i.e. they are uncoded). The accusative and the oblique case are used to code definite direct objects; the former is usually found in the nominative alignment, the latter is more common in the ergative, even though it is by no means rare to find definite direct objects in the oblique case in the present domain or the accusative in the past domain (Korn 2008: 256). This is illustrated in (6.53a, b). In the past domain, where the ergative construction applies, it is possible to have both agent and definite direct object in the oblique case, as in (6.53c):

(6.53) Balochi (Indo-European, Indo-Iranian)

- (a) *man čuk-ā ras-ēn-ā*  
 1SG child-PL.OBL/DOM arrive.PRS-CAUS-1SG  
 "I transport the children" (Jahani and Korn 2009: 659)
- (b) *kučik-ā hamā ĵīnik-ārā dīst*  
 dog-OBL that\_very girl-DOM see.PST  
 "The dog saw this girl" (Southern Balochi; Korn 2008: 261)
- (c) *bačakk-ā watī dantān-ā prōšt*  
 boy-OBL own tooth-OBL.PL break.PST  
 "The boy broke his teeth" (West Balochi; Korn 2008: 260)

Balochi thus presents a rather complex system in which DOM can be expressed by means of two case markers, the oblique and the accusative, which at first glance do not show any complementary distribution. However, as Korn (2008) has argued, the data shows that it is possible to make a preliminary distinction between the usages of the two case markers. Based on her data, it appears that DOM via oblique case marking is primarily influenced by definiteness, while the accusative is restricted to human definite direct objects (also in the present). However, as observed by Korn herself (2008: 263), the picture seems to be much more intricate, as there are many cases in which both definite and human definite direct objects are left uncoded. More data is needed to work out the exact conditions of appearance of DOM. It seems to be likely, however, that information-structure

parameters are crucial in determining the presence of DOM, as observed by Farrell (1995: 224), who asserts that DOM is used when "specific emphasis" on the direct object is needed.

A similar system is attested in Vafsi (Stilo 2010), which has a DOM system based on saliency, operationalised in terms of animacy/definiteness. Unlike Balochi, however, Vafsi employs the same marker, i.e. the oblique, for DOM both in the present (nominative alignment) and in the past (ergative alignment). As a consequence, in the past both arguments of a transitive clause may be coded in the oblique (Stilo 2010: 246).

Interestingly, Romani dialects (Matras 2002) display a similar connection between topicality and DOM. As pointed out by Matras (2002: 82), Romani DOM seems to be a case of preservation rather than innovation, and it is usually employed with topical direct objects. As a matter of fact, DOM (expressed by the so-called "independent oblique") is nowadays obligatory with pronouns as a result of grammaticalisation, while it shows a great amount of variation within the domain of animates, with definite animate direct objects being more likely to receive DOM than indefinite non-specific ones (Matras 2002: 86). Matras further argues that topicality is the common factor that underlies all the uses of the independent oblique marker, which is also employed to code possessors and datives (Matras 2002: 86-87). In his view, the opposition between zero-coding (i.e. nominative) and overt coding is an opposition between non-topical vs. topical status of the overtly coded constituent. Of course, as Matras himself observes, this topic coding function is variably grammaticalised in Romani, as the obligatory presence of DOM with pronouns suggests.

Let us now turn to the Dravidian language Badaga (Pilot-Rachoor 1994). At first glance, DOM in Badaga seems to be conditioned by the humanness and definiteness of the direct object. As a matter of fact, human nouns, as well as pronouns and NPs introduced by a deictic expression or modified by a participial construction are usually overtly coded (as in 6.54a, c; Pilot-Rachoor 1994: 364-367). It is possible, however, that a human NP is left uncoded, as in (6.54b):

(6.54) Badaga (Dravidian, Southern-Dravidian)

(a) *ama ondu manusa-na nooḍida*  
 3SG one man-DOM see.3SG.PST  
 "He saw a man"

(b) *avve kuusu origiciya*  
 mother child sleep.CAUS.PST.3SG.F  
 "The mother sent the child to sleep" (Pilot-Rachoor 1994: 364)

(c) *naa ii ondu kaaqida-oa oodide*  
 1SG this one book-DOM read.PST.1SG

"I've read one of these books" (Pilot-Rachoor 1994: 364)

As Pilot-Rachoor (1994) rightly observes, differences in information-structure play a role in the presence of DOM. Specific direct objects may be overtly coded, as the example (6.54c) shows, when they are prominent in the discourse. In addition, Badaga (word order SOV) shows the tendency to overtly code topical dislocated direct objects, as can be seen in (6.55a, b):

(6.55) Badaga (Dravidian, Southern-Dravidian)

(a) *ii niir-a kuusu susi buiia*  
 this water-DOM child spill.PTCP.PST PFV.3SG  
 "The water, the child spilt it"

(b) *enna aṇṇa-na kunni kaccida*  
 my elder\_brother-DOM bee sting.PST.3SG  
 "My brother, a bee stung him"

(c) *bukk-a obba oodina*  
 book\_DOM someone read.PRS.3SG.M  
 "As for books, someone reads them" (Pilot-Rachoor 1994: 386)

Differences in topicality can account for the fact that abstract nouns, which are at the low end of the definiteness hierarchy, can be overtly coded, as in (6.56):

(6.56) Badaga (Dravidian, Southern-Dravidian)

*ava (...) tanna nambike-ya budule*  
 3SG.F (...) POSS.3SG.F confidence-DOM relax.NEG  
 "She has not lost her courage" (Pilot-Rachoor 1994: 370)

## 6.6 Altaic

In this section I will examine DOM in Altaic languages. As will be shown below, the parameters governing the presence of case marking or indexation with direct objects are definiteness/specificity and topicality, although there is a certain amount of variation which can be accounted for in terms of grammaticalisation.

Turkish is one the most studied languages with regard to DOM (see Enç 1991, Erguvanli 1984, Kornfilt 1997, von Heusinger and Kornfilt 2005, among others). Traditionally, DOM has been taken as an indicator of specificity. This is shown by examples (6.57): in (6.57a, b), the direct object receives overt accusative marking insofar as it is definite and indefinite specific respectively. When indefinite non-specific, it stays uncoded, as in (6.57c):



## (6.57) Turkish (Altaic, Turkic)

- (a) *(Ben) kitab-ı oku-du-m*  
 1SG book-DOM read-PST-1SG  
 "I read the book"
- (b) *(Ben) bir kitab-ı oku-du-m*  
 1SG a book-DOM read-PST-1SG  
 "I read a certain book"
- (c) *(Ben) bir kitap oku-du-m.*  
 1SG a book read-PST-1SG  
 "I read a book" (von Heusinger and Kaiser 2005: 8)

As noted by von Heusinger and Kornfilt (2005: 10), sometimes a generic direct object may be overtly coded. One of the most interesting property of DOM in Turkish is its sensibility to word order, a fact that I have noticed in other languages as well. Turkish basic word order is SOV, but other permutations are possible based on information structure. For instance, the sentence-initial position is exploited to topicalise a constituent (Kornfilt 1997: 200 ff.). Uncoded direct object, however, must occupy the preverbal position and cannot be moved (Kornfilt 1997: 215). When a direct object is topicalised (i.e. it is put in sentence-initial position<sup>12</sup>, it must be overtly coded, as shown by (6.58):

## (6.58) Turkish (Altaic, Turkic)

- (a) *uskumru-yu Hasan piş-ir-id, istakoz-u da Ali*  
 mackerel-DOM Hasan cook-CAUS-PST lobster-DOM and Ali  
*ye-di*  
 eat-PST  
 "As for the mackerel, Hasan ate (it), and as for the lobster, Ali ate (it)" (Kornfilt 1997: 205)
- (b) *Mavi kaplı bir kitab-ı Murat aceleyle oku-yor.*  
 blue covered a book-DOM Murat hurriedly read-PRS.PROGR  
 "Murat is hurriedly reading a (certain) blue-covered book"
- (c) *Bir öğrenci-yi arı-yor-um.*  
 a student-DOM look\_for-PRS.PROGR-1SG  
*Bul-a-mı-yor-um.*  
 find-NEG.ABIL-NEG-PRS.PROGR-1SG  
 "I am looking for a student. I can't find him (\*one) (von Heusinger and Kornfilt 2005: 12)

<sup>12</sup>It must be noted that Turkish topicalisations/dislocations do not involve the presence of a resumptive element (Kornfilt 1997: 205).

As von Heusinger and Kaiser (2005: 12-13) rightly observe, the specificity constraint is due to the fact that indefinite non-specific elements are usually disallowed in topical positions. As a matter of fact, these positions usually force specific interpretations, as is demonstrated by the pronominal anaphora, which must be "him" in (6.58c). Thus, in Turkish too, there is a strong correlation between dislocation/topical position and presence of DOM. There is a further semantico-pragmatic parameter, namely specificity, which interacts with DOM. In my view, this is easily accounted for by the fact that indefinite non-specific referents cannot be construed as topics.

Likewise, in other Altaic languages, DOM is strongly dependent upon information-structure factors. In Mangghuer (word order SOV; Slater 2003), DOM is used with definite and indefinite specific direct objects, as in (6.59b) as opposed to (6.59a):

(6.59) Mangghuer (Altaic, Mongolic)

- (a) *gan mori wuni-jiang*  
3SG horse ride-OBJECT.PFV  
"S/he rode a horse, went horse-riding"
- (b) *gan mori-ni wuni-jiang*  
3SG horse-DOM ride-OBJECT.PFV  
"S/he rode the horse" (Slater 2003: 165)

Again, DOM becomes obligatory when a direct object is topicalised in sentence-initial position (Slater 2003: 189 ff.), provided that it is definite and/or specific, as shown by the examples in (6.60)<sup>13</sup>:

(6.60) Mangghuer (Altaic, Mongolic)

- (a) *Dimeini bi he-ji xi-a bai*  
bread-DOM 1SG take-IPFV go-VOLUNT EMPH  
"The bread, let me take (it)"
- (b) *gan mori-ni wuni-jiang*  
3SG horse-DOM ride-OBJECT.PFV  
"S/he rode the horse" (Slater 2003: 165)
- (c) *Gaga-ni gan-si lake gher-gha-jiang*  
elder\_brother-DOM 3SG-PL pull go.out-CAUS-OBJECT.PFV  
"Elder Brother, they dragged out" (Slater 2003: 124)

<sup>13</sup>Interestingly, there seems to be a topic-shift function in the use of DOM with dislocated direct object. Example (6.60a), taken from the "Rabbit's Trick" story, signal a shift from one topic (the drum), to another (the bread). See Slater (2003: 343) for the whole text.

A similar situation is found in Bao'an Tu (Fried 2011), where DOM is restricted to definite direct objects, as shown by (6.61a) vs. (6.61b):

(6.61) Bao'an (Altaic, Mongolic)

(a) *aku-gə*                      *ɛazə-nə*      *jiχ-tɕo*  
 woman-SG.INDEF child-DOM hit-IPFV.OBJECT  
 "A woman hit the child"

(b) *gaku-gə*                      *ɛazə*      *jiχ-tɕo*  
 woman-SG.INDEF child hit-IPFV.OBJECT  
 "A woman hit a child/children" (Fried 2011: 47-48)

The enclitic *-nə* marks possessors and some temporal expressions such as dates (Fried 2011: 50-51). As is well-known, possessors and temporal expressions are often topical in discourse. The close link of DOM and topicality is confirmed by Bao'an too, where topicalised direct objects receive DOM (Fried 2011: 255 ff.), as shown by (6.62a, b):

(6.62) Bao'an (Altaic, Mongolic)

(a) *wadz-ə*      *χani-la*      *tɕ<sup>h</sup>ajum-kə-tɕə*                      *mər*      *k<sup>h</sup>ər-na*  
 sock-DOM all-PL collective-VBZ-IPFV wear be\_required-DUR  
 "As for socks, everyone should wear the same (kind)"

(b) *atɕaŋ-nə*      *hkudə*      *Dortɕə*      *jiχ-saŋ*  
 3SG-DOM yesterday Dorje hit-POS  
 "Him, Dorje hit yesterday" (Fried 2011: 256)

DOM in Manchu (Tungusic; Gorelova 2002: 167 ff.) seems to behave in a very similar way. Besides coding definite and specific direct objects, DOM is obligatorily found with topicalised direct objects (i.e. in OSV constructions). Likewise, in Udihe (Tungusic), DOM is generally absent with non-specific direct objects and specific new direct objects, i.e. with object introduced in the text for the first time (Nikolaeva and Tolskaya 2001: 119 ff.). When the direct object is topicalised by means of dislocation, DOM becomes obligatory even when its non-topicalised counterpart is not (Nikolaeva and Tolskaya 2001: 120). A similar constraint is found in Khalka-Mongolian, where topicalised direct objects must be overtly coded (Guntsetseg 2008):

(6.63) Mongolian (Altaic, Mongolic)

*neg zahia-g bi bich-sen*  
 a letter-DOM 1SG write-PST  
 "A letter, I wrote"

Geist, Guntsetseg and von Heusinger (2010) have indeed shown that DOM in Mongolian is one of the devices used to promote a referent to topic in the following discourse.

Likewise, in Altai DOM is reserved to topical direct objects. Altai has a grammaticalised hearsay particle *dežet* "they say", where *de-* is the grammaticalised form of the verbum dicendi (Skribnik 2001b: 351). Albeit grammaticalised, this particle still maintains the governing pattern of a transitive verb. Consequently, the direct object can be overtly coded or uncoded depending upon its topical/focal status. Examples (6.64a) and 6.64b) nicely illustrate the topicality requirement for the presence of DOM. (6.64a) is the reply to the question "And who is the best student in their class?", where *Aržana* is left uncoded as it constitutes the focal portion of the sentence. By contrast, (6.64b) is a reply to the question "How is Aržana?", where *Aržana* is the presupposed, topical element. As expected, the direct object in (6.64b) must be overtly coded:

(6.64) Altai (Altaic, Turkic)

- (a) *Aržana d'akši üren-ip tur-gan dežet*  
 Aržana well learn-CONV AUX-PFV.3SG HSY  
 "They say that Aržana studies well"
- (b) *Aržana-ni d'akši üren-ip tur-gan dežet*  
 Aržana-DOM well learn-CONV AUX-PFV.3SG HSY  
 "They say that Aržana studies well" (Skribnik 2001b: 352)

In other Altaic languages, such as Kirghiz (Turkic; Imart 1981), Tuvan (Turkic; Harrison 2001) and Uzbek (Ismatullaev and Feldman 1995), DOM is obligatory with NPs in the high end of the definiteness hierarchy (thus conforming to the situation found in the other Altaic languages examined so far), while it is optional with indefinite specific NPs. I hypothesise that the obligatoriness of DOM with NPs high on the definiteness hierarchy is a result of the conventionalisation of DOM to definite direct objects, given that definiteness is a typical feature of topics. As a matter of fact, Skribnik (2001b: 352-353) demonstrates that the correlation between DOM and topicality still holds in many Altaic languages, as in Buriat, where DOM is now obligatory with pronominal and human direct objects (see also Poppe 1960).

## 6.7 Uralic

In this section I will examine the distribution of DOI and DOM among Uralic languages. As is well known in Uralistic literature, several Uralic languages (belonging to the Ugric and Samoyedic branches) show a twofold verbal inflectional paradigm, i.e. the so-called "subjective" and "objective" conjugation, the latter representing DOI. The use of one or the other conjugation has traditionally been explained in terms of (in)definiteness of the direct object, with the objective conjugation used when the direct object is definite and the subjective conjugation when it is not (Abondolo 1998). However, the rule is not as hard and fast as it may seem. A growing body of literature has identified topicality as the main feature that drives the usage of one form over the other (Nikolaeva 1999, 2001; Skribnik 2001a, among others). In addition, some languages show DOM too. I will begin by analysing a language that has both strategies, i.e. Eastern Mansi (Virtanen submitted) and the languages that show DOI alone (i.e. Ostyak, Northern Mansi and Hungarian). Finally, the interesting situation found in Komi will be discussed.

Eastern Mansi still retains the accusative case, unlike Northern Mansi (see below) and has DOI. Thus, it is possible to have DOI only (6.65a), DOM and DOI simultaneously (6.65b), and only DOM (6.65c; see Virtanen submitted: 8):

(6.65) Eastern Mansi (Uralic, Ugric)

- (a) *towøl.wojøl jål-pøšøwl-øš-tø*  
 then            PREF-blow-PRET.SUBJ.SG-OBJ.3SG  
 "Then he blew it (i.e. the fire) out"
- (b) *jiiw-ty-mø sok juw-tåtø-s-te*  
 wood-PL-DOM all home\_bring-PRET.SUBJ.SG-OBJ.3SG  
 "He brought all the wood home"
- (c) *näär sons-i, sågrøp-mø sons-i*  
 what look-PRS.3SG axe-DOM look-PRS.3SG  
 "Wherever he looks, he sees the axe" (Virtanen submitted)

The Eastern Mansi examples are taken from a corpus of texts (Virtanen submitted). Virtanen demonstrates that, although all the three strategies share topicality as a common denominator, there seems to be significant differences in their distribution. Whilst DOI alone is reserved to most topical direct objects (being thus comparable to a zero anaphora in Givón's topic accessibility scale), DOM and DOI co-occur when the direct object shows a lower degree of topicality which makes it necessary to refer to it via the full NP. DOM alone signals a contrastive topic that serves to shift the current topic of discourse from one item to another. Finally, the so-called subjective conjugation (i.e. absence of DOI) is found with

new, focal direct objects. Eastern Mansi therefore abides by my hypothesis that topicality is the common factor behind the presence of both DOM and DOI, but to a different extent. While DOI is primarily reserved to the most topical direct objects, DOM appears with less-continuous referents, being mainly used in topic-shift structures.

Unlike Eastern Mansi, Northern Mansi (Skribnik 2001a) shows DOI only, but patterns with Eastern Mansi as to the main governing factor, i.e. topicality. As Skribnik points out, the presence of DOI positively correlates with the topicality of the direct object, while its absence indicates that the direct object is focal. Skribnik's analysis of Northern Mansi -based upon a corpus of 2200 clauses from tales and traditional stories- appeals to the concepts of primary and secondary topics put forward by Nikolaeva (1999, 2001) for DOI in Northern Ostyak which I have discussed in Chapter 2.

Recall that DOI in Northern Ostyak is not obligatory, but depends upon the topical status of the direct object. Nikolaeva (2001) elegantly accounts for the absence of DOI in (6.66a) and its presence in (6.66b) in terms of topicality status. The direct object of the (6.66a) is not topical, and therefore it does not trigger DOI. The direct object in (6.66b), however, is a secondary topic, and its topical status triggers DOI.

(6.66) Northern Ostyak (Uralic, Ugric)

- (a) *ma tam kala we:l-s-əm*  
 I this reindeer kill-PST-1SG.SUBJ  
 "I killed this reindeer"
- (b) *ma tam kalan we:l-se-nil-am*  
 I this reindeer kill-PST-DU.OBJ-1SG.SUBJ  
 "I killed these (two) reindeer" (Nikolaeva 2001: 16)

Her hypothesis is further corroborated by the distribution of DOI absent with *wh*-question words, as well as the answers to these questions, which are usually linked to focality. As a matter of fact, the presence of DOI results in the ungrammaticality of the clause, as shown by examples (6.67a) and (6.67b):

(6.67) Northern Ostyak (Uralic, Ugric)

- (a) *mati kalan we:l-es/\*we:l-s-elli*  
 which reindeer kill-PST.3SG.SUBJ/kill-PST-OBJ.3SG.SUBJ  
 "Which reindeer did he kill?"
- (b) *tam kalan we:l-es/\*we:l-s-elli*  
 this reindeer kill-PST.3SG.SUBJ/kill-PST-OBJ.3SG.SUBJ  
 "He killed this reindeer" (Nikolaeva 2001: 17)

As discussed by Dalrymple and Nikolaeva's (2011: 127 ff.) a similar pattern is found in Tundra Nenets, where only secondary topic direct objects trigger DOI. Interestingly, though, in Tundra Nenets DOI is prohibited when the direct object is a first/second person pronoun (both singular and plural), as shown by Dalrymple and Nikolaeva (2011). A similar constraint is attested across the Samoyedic languages Enet, Nenets, Nganasan, and Selkup, which have DOI only with a third person direct object (Körtvély 2005). Eastern Ostyak, one of the Ostyak dialects, has the same restriction on DOI, unlike Northern Ostyak (Gulya 1966: 104). Like Tundra Nenets, no DOI is permitted when the direct object is a first/second person pronoun.

The most well-known example of DOI limited to third person direct objects is Hungarian. DOI is present when the verb governs a third personal pronoun (overt or null, 6.68a), a proper name (6.68b), a noun accompanied by a definite determiner (6.68c) or by a possessive suffix (6.68d), and a direct object clause (Rounds 2001 16-18; Coppock and Wechsler to appear)<sup>14</sup>:

(6.68) Hungarian (Uralic, Ugric)

- (a) *Lát-om*  
see-1SG.OBJ  
"I see him/it"
- (b) *Lát-om Zsuzsá-t*  
see-1SG.OBJ Zsuzsa-ACC  
"I see Zsuzsa" (Rounds 2001: 17)
- (c) *Eltitkol-om valamennyi találkozás-t*  
keep.secret-1SG.OBJ each meeting-ACC  
"I keep each meeting secret"
- (d) *Olvas-om Péter vers-é-t*  
read-1SG.OBJ Peter poem-3SG.POSS-ACC  
"I am reading Peter's poem" (Coppock and Wechsler to appear)

DOI in Hungarian is triggered by the definiteness of the direct object, as has been repeatedly observed in the literature (Moravcsik 1974, Lazard 2001, É. Kiss 2011, Coppock and Wechsler to appear, among others). DOI is indeed absent (i.e. the so-called subjective conjugation is used) when the direct object is indefinite or is a first/second person pronoun, just like in the Samoyedic languages discussed

<sup>14</sup>It is remarkable that when the agent is first person singular and the direct object is second person, the special form *-lak/lek* must be used (Rounds 2001: 18). Furthermore, DOI is used when the direct object is a reciprocal or reflexive pronoun. See Coppock and Wechsler (to appear) for a thorough discussion of these cases.

above. Unlike Samoyedic languages, however, Hungarian DOI does not hinge upon information structure at all, as DOI is brought about by the formal definiteness of the direct object and can be found with both topical and focal direct objects (Coppock and Wechsler to appear).

Nonetheless, as proposed by Marcantonio (1985) the definiteness constraint at work in Modern Hungarian DOI is the result of the extension of DOI from topical direct objects to direct objects showing a feature usually associated with topicality, namely definiteness. Indeed, Marcantonio (1985), based on evidence from Ostyak and Mansi, as well as data from Old Hungarian, showed that DOI in Hungarian initially emerged in OSV constructions to encode the topic function of the direct object and was not influenced by its definiteness. Furthermore, in Marcantonio's (1981: 229) reconstruction, the Hungarian accusative marker, which is now obligatory for every type of direct object, was initially used to encode topicalised direct objects and was later extended to every direct object, similarly to what happened with DOI. Therefore, Marcantonio's analysis complies with my hypothesis that, diachronically, DOM arises in dislocated constructions.

Uralic languages, however, show quite a peculiar pattern, since the absence of DOI with first/second person direct objects does not conform to the predictions of the animacy/definiteness hierarchies. To account for this apparently aberrant situation, Dalrymple and Nikolaeva's (2011: 194-201) propose that the DOI in Proto-(Eastern)-Uralic was found with all topical direct objects, and later restricted to third person topical direct objects because of the recurrent connection between third person and (secondary) topicality. By contrast, topical first/second person pronouns are more likely to be primary topics and thus show up as subjects. Therefore, they conclude that Samoyedic languages and Hungarian have grammaticalised the tendency for third persons to be secondary topics as opposed to first/second person pronouns, which are more likely to be primary topics. For instance, in Ostyak, first/second person pronouns in direct object position require passivisation in order to match the information structure roles<sup>15</sup>.

As I have already discussed, while Nikolaeva (2001) and Dalrymple and Nikolaeva's (2011) view of DOM and DOI as secondary topics accounts for the distribution of DOI in Ugric languages, yet the notion of secondary topic does not succeed to capture the link of DOM with dislocated direct objects, which are primary topics.

Let us now turn to Komi. Komi exhibits quite a complex DOM system. Direct objects can be overtly coded in different ways. Direct objects are overtly coded with the accusative when they are constituted by NPs high on the animacy/definiteness hierarchies, i.e. personal pronouns (6.69a), proper names, NPs

<sup>15</sup>See Siewierska (2004:150-151) for a list of languages that show indexation with the third person only.



accompanied by a demonstrative or a possessive suffix indicating possession and givenness (6.69b) (Klumpp 2009; to appear). The accusative has three forms, two (-*tö*, -*sö*) used when the direct object is modified by the possessive suffix and one (-*ös*) reserved to non-possessed (mostly) human direct objects (6.69c; Klumpp to appear):

## (6.69) Komi (Uralic, Permic)

- (a) *Korkö najö kor-isny sud vyl-ö*  
 once they.ACC call-PRET.3PL court on-ILL  
 "One day, they were summoned to court"
- (b) *Aski-nas Vasa vajas kujim t'svet.*  
 other\_day-INS.3SG water.ADJ bring.PRS.3SG three flower  
*Sar-pi med mi'tsa t'svet-s-ö boštas*  
 tsar-son SUP beautiful flower-POSS.3SG-ACC take.PRS.3SG  
 "The other day the waterghost brings three flowers. The tsarevich  
 takes the most beautiful flower"
- (c) *Najö sjiö d'ivö ad'd'žasny i kaga-ös boštasny i*  
 they this wonder see.3PL and child-ACC.AN take.3PL and  
*vod'žö munasny*  
 further go.3PL  
 "They see this wonder and pick up the child and continue their way"  
 (Klumpp to appear)

Thus, the pattern surveyed so far is a clear case of DOM governed by topic-worthiness features like identifiability and animacy. No information structure distinctions condition the distribution of DOM in Komi. In fact, as demonstrated by Klumpp (to appear), in the dialects of Vym' and Izhma, a third strategy has emerged to overtly code differences in the information structure status of the direct object. The dative case-marker may be used with direct objects that are topical, as the following example shows:

## (6.70) Komi (Uralic, Permic)

- Ivan menö vajan, on? Vot ješli-kö te vajan*  
 Ivan me.ACC bring.2SG NEG.2SG PT if-CP you bring.2SG  
*menö-ly, te asy lok tat't's-ö*  
 me.ACC-DOM you tomorrow come.IMP.2SG here-ILL  
 "Ivan, [will you MARRY me]<sub>FOC</sub>, or won't you? So if you then marry  
 [me]<sub>TOP</sub>, come here tomorrow" (Klumpp to appear)

Possessed DO	Non-possessed Human DO	Topical DO
accusative <i>-tö, sö</i>	accusative <i>-ös</i>	dative <i>-ly</i>

Table 6.2: Distribution of DOM in Komi dialects

The different forms used to overtly code direct objects in Komi are summarised in table (6.7):

In (6.70) we find the first person direct object "me" first in its default accusative form (*menö*), and then coded by the dative (*menöly*). As Klumpp argues based on corpus data, this morpho-syntactic contrast is caused by the different information status of the direct object pronoun. The accusative coded form is focal, while the dative coded form (where the dative follows the morphological accusative) is topical, as demonstrated also by the fact that the dative coded direct object is de-accented (Klumpp to appear). The link of DOM and topicality thus holds again. It seems to be probable that, diachronically, DOM in Komi (i.e. the zero vs. accusative) was based upon information structural properties too, and then started to be extended to every topic-worthy direct object irrespective of its actual topicality. The accusative forms are found with every direct object that comply with the right semantic conditions. Thus, the cycle has started again: the overt coding of topical direct objects was taken over by the dative (although this is dialectally restricted). Meanwhile, the other forms became generalised to focal objects too and then have now started to be used as focus markers.

## 6.8 Yukaghir

Kolyma Yukaghir (Maslova 2003b) has an elaborated system of object marking, which works differently depending on the finite vs. non-finite status of the clause. There are four different options to encode direct objects in finite clauses, as shown in Table (6.3):

	1st/2nd	3rd definite	3rd indefinite
DO if A 1/2	pron acc		nominative
Focus	predicative	nominative	predicative
DO if A is 3		accusative	accusative/instrumental

Table 6.3: Object encoding in Kolyma Yukaghir (adapted from Maslova (2003b: 89))

Thus, a third person direct object will be in the nominative case (i.e. zero) if the agent is first or second person (6.71a). First and second direct objects instead are encoded via the special pronominal accusative *-ul* (6.71b). If the agent is third person, direct objects can be overtly coded either via the accusative or the instrumental. Tendentially, definite direct objects (such as pronouns, proper names, possessive NPs, and all definite direct objects) receive the accusative (6.71c), while indefinite direct objects receive the instrumental (6.71d). Finally, direct objects in focus receive what Maslova (2003b: 91) calls "predicative" case (here glossed as "focus"). Again, the predicative case is incompatible with possessive NPs, proper names and third person pronouns, which are encoded via the nominative when focal (6.71e):

(6.71) Kolyma Yukaghir (Yukaghir, Southern)

- (a) *met mēmē ijī*  
 1SG.NOM bear be.afraid.TR.1SG  
 "I am afraid of the bear"
- (b) *met-ul amde-l-get polde-mek*  
 1SG-PRON.ACC die-PFV-ANR-ABL save-TR.2SG  
 "You saved me from death"
- (c) *tet kimnī met-kele kudede-m*  
 your whip me-ACC/DOM kill-TR.3SG  
 "Your whip has killed me"
- (d) *n'umud'i-le mid'-u-m*  
 ax-INSTR take-0-TR.3SG  
 "He took an ax"
- (e) *met tet-in šaqale-lek kej-te-me*  
 1SG.NOM 2SG-DAT fox-FOC give-FUT-OBJ.FOC.1SG  
 "I will give you a fox" (Maslova 2003b: 89-95)

The variation in object encoding found in finite clauses is primarily regulated by the person of the agent. Nonetheless, the connection between topicality and accusative case is still observable, since the accusative is the only possible option with definite NPs, which receive nominative encoding if focal.

In non-finite clauses, the situation is slightly different. First of all, the predicative (focus) case is not allowed. As in finite clauses, the nominative is the only possible option available when the agent is first/second person. Unlike finite clauses, when the agent is third person, the direct object can be either overtly coded through the accusative or zero-coded (i.e. nominative). According to Maslova

(2003b: 331 ff.), the two different options in the encoding of direct objects with third person agents are motivated by differences in information structure, since zero-coding can occur with definite NPs as well. In Maslova's view, the presence of overt coding depends upon the prominence (i.e. topicality) of the direct object at the text level (based on the hierarchy clause < episode < text). When the agent is the more topical entity, the direct object will be zero-coded. By contrast, when the direct object is more topical than the agent, it will be in the accusative, as shown by (6.72a) as opposed to (6.72b):

(6.72) Kolyma Yukaghir (Yukaghir, Southern)

- (a) *tudel tude čohojo min-delle šā-n kiel-gen jaqlude*  
 3SG his knife take-SS:PFV tree-ATTR chink-PROL further  
*čolha-j-m*  
 push-PFV-TR:3SG  
 "He took his knife and poked it through a chink in the wood"
- (b) *tintaj šøgī-gele el-l'uø-čuøñ paja-j-m*  
 that bag-ACC/DOM NEG-see-PRV carry-PFV-TR.3SG  
 "Without having a look at this bag, he put on his back" (Maslova 2003b: 335)

The direct object in (6.72a) is not a prominent entity in the story, and thus stays uncoded. By contrast, the direct object in (6.72b) plays a significant role in the text and is thus overtly coded. Thus, direct object encoding in non-finite clauses in Kolyma Yukaghir is primarily motivated by the topicality of the direct object, as opposed to the pattern found in finite clauses, where DOM relies primarily upon the person of the agent. This latter pattern could be interpreted as the grammaticalisation of frequent correlations between discourse participants, their topicality and their likelihood to appear in a certain role in grammar. First and second person participants are usually the most topical ones, and tend to appear as agents. If the direct object is topical, or more topical than the agent, it receives overt coding.

## 6.9 South America

In this section I provide an overview of DOM and DOI in a sample of languages from South America. As will be thoroughly discussed below, DOM and DOI systems in South American languages lend further support to the view that it is topicality the main factor underlying the appearance of DOM and DOI. Indeed, as observed in Zúñiga (2007), the distribution of DOM in many languages of South

America is hard to account for if information-structural parameters are not taken into consideration in the analysis of such systems. The data that will be presented in the following strongly underpins the role of topicality in shaping DOM and, more generally, case marking.

The Tucanoan languages in the sample (Desano, Retuarã, Cubeo, Tucano, Barasano, Koreguaje, and Siona) employ the suffix *-re* and various allomorphs to overtly code direct objects (Zúñiga 2007: 216). The details about the use of the object marker in each language vary, but all of them seem to have grammaticalised DOM with the most topic-worthy NP categories, i.e. pronouns and proper names. The following examples are from Tucano (Zúñiga 2007)

(6.73) Tucano (Tucanoan, Eastern-Tucanoan)

- (a) *Bdĩ-de ko'té-nĩ-a-pĩ*  
 2SG-DOM wait-FRUSTR-REC.PST-VIS.SAP  
 "I waited for you, in vain"
- (b) *Peduru-de paâ-gĩ' weé-'*  
 Peduru-DOM hit-IMPL.SS AUX-VIS.SAP  
 "I am hitting Peduru" (Zúñiga 2007)

In other Tucanoan languages, DOM has been extended along the animacy hierarchy, as in Retuarã (Strom 1992: 59 ff.), Cubeo (Morse and Maxwell 1999: 110 ff.), Koreguaje (Cook and Criswell 1993: 48), and Kotiria (Stenzel 2008). Siona is the only one that has obligatory DOM with pronouns only (Wheeler 1970). A similar situation is found in Makú languages (see Chapter 5), where DOM has been extended to human direct objects, with the exception of Dâw (Martins and Martins 1999), where DOM is still connected to the high topicality of the direct object.

By contrast, Barasano (Jones and Jones 1991: 65 ff.) and Desano (Miller 1999: 57) have extended DOM to definite direct objects, with animacy playing no role in the presence of the object marker, as shown by the following Barasano examples (6.74a, b):

(6.74) Barasano (Tucanoan, Eastern Tucanoan)

- (a) *B#-re sĩa-g#-bĩ yai*  
 2SG-DOM kill-SG.M-3SG.M wildcat  
 "The wildcat will probably kill you"
- (b) *rase-a sĩa-to bãdi*  
 toucan kill-IMP.2 1PL.INCL  
 "Let's kill toucans" (Barasano; Jones and Jones 1991: 65)

The animacy/definiteness constraints, however, are not so rigid. As a matter of fact, it is possible, e.g. in Cubeo, to find DOM with inanimate direct objects, provided that it is not sentence-final and not newly introduced into the discourse (Morse and Maxwell 1999). All the authors of the grammars of Tucanoan languages clearly state that the main requirement for DOM to appear is that the direct object be topical. Apart from the analysis of the distribution of DOM in discourse, the link with topicality is also corroborated by the fact that the object marker in all the languages examined here is used also to indicate spatial and temporal expressions, as shown by the following examples from Tucano:

## (6.75) Tucano (Tucanoan, Eastern Tucanoan)

- (a) *a'to-de, dō'ó-pi kãdi-gi dñi-a-ti?*  
 here-DOM where-FOC sleep-NMLRZ.SG.F be-REC.PST-VIS.Q  
 "Here (i.e. in the city), where do you sleep?" (Zúñiga 2007: 218)
- (b) *Dĩ'káa-de, bu'ê-dã!*  
 today-DOM study-IMP  
 "Today, let's study!" (Zúñiga 2007: 218)

Following Ramirez (1997), Zúñiga (2007) proposes that the possible link between the use of the same marker both on direct objects and datives is the meaning of *-re* as "about, with respect to" (Zúñiga 2007: 218), as in (6.76a) and (6.76b):

## (6.76) Tucano (Tucanoan, Eastern Tucanoan)

- (a) *Dũbño-de uúkũ-a-bã*  
 woman-DOM talk-REC.PST-VIS.3PL  
 "They talked about the woman"
- (b) *Yahá-'ke-de bĩ'i yẽ'édõho*  
 steal-NMLZR.INAN.P.PFV-DOM 2SG what  
*bãsi-sa-di?*  
 know-PRES.NVIS-Q  
 "What do you know about the theft?" (Zúñiga 2007: 218)

Moreover, quoting Ramirez (1997), Zúñiga (2007: 218) argues that the "use of *-re* with STEs (spatial and temporal expressions, GI) is an extension of its basic function: the counterpart of highly individuated, referential objects is the topicalization or discursive salience of STEs" (Zúñiga (2007)). In these languages too, then, the polysemy of the object marker upholds the link with topicality. A further proof of the close connection of DOM and topicality comes from Koguje (Cook and Levinsohn 1985), Tucano (Ramirez 1997), and Desano (Miller

1999). As I have argued above, information structure strongly affects the distribution of case-marking in all Tucanoan languages. Desano, Tucano, and Koreguaje are somewhat more extreme in mapping grammatical relations with information-structure roles. Topical direct objects in Desano are overtly coded by means of *-re*, while focal subjects/agents are overtly coded by means of the suffix *-pi* (Miller 1999: 161-162). A similar pattern is attested in Tucano (Ramirez 1997: 231) as well. Koreguaje is even more extreme in reflecting information structure notions into core argument marking. Aside from overtly coding topical direct objects, Koreguaje has two different markers for focused animate direct objects (*-ni*) and focused subjects/agents (*-pi*) as well as inanimate direct objects (Cook and Levinsohn 1985). The presence of overt coding with focal animate direct objects is probably explained by the tendency for animate direct objects to occur as topical elements rather than focal ones in discourse. Examples (6.77a, b) show the distribution of the markers. The agent of (6.77a) is topical and then uncoded; in contrast, in (6.77b), the focal agent is overtly marked, as is the topical direct object. Example (6.77c) shows the use of the animate focus marker *-ni* on a new participant, which is subsequently mentioned again in direct object position. In the second occurrence, the topical direct object is coded via the topical object marker:

(6.77) Koreguaje (Tucanoan, Western Tucanoan)

- (a) *wanisōasomɤ campesino*  
kill.PST.3SG.M country\_fellow  
"The country-fellow killed (him)"
- (b) *mɤ'ɤ-pi chɤ'ɤcho'ojeɤ'-te huanisōmɤ*  
you my.younger.brother-DOM kill.PST.2SG.M  
"[You-FOC]<sub>FOC</sub> killed [my younger brother!]<sub>TOP</sub>" (Cook and Levinsohn 1985: 93)
- (c) *pa'iɤ Ru-ni José cahuɤna repaɤ'-te r̥iso huēasōra*  
being Ru-AN.FOC José neck 3SG-DOM strangle kill  
*chiniasomɤ*  
want.PST.3SG.M  
"Then José tried to strangle [Ru]<sub>FOC</sub> and kill [him]<sub>TOP</sub>" (Cook and Levinsohn 1985: 93)

Table (6.9) summarises the distribution of case marking in Desano, Tucano, and Koreguaje:

A system similar to the ones found in Desano and Tucano is found in Tari-ana, an Arawak language spoken in the Vaupés region of Amazonia (Aikhenvald

Language	Topical DO	Focal S/A	Focal Animate DO
Desano, Tucano	<i>-re</i>	<i>-pi</i>	∅
Koreguaje	<i>-re</i>	<i>-pi</i>	<i>-ni</i>

Table 6.4: Distribution of case marking based on information structure in Tucanoan languages

2003). Indeed, many scholars have noticed that Tariana would have developed its case-marking system due to contact with Tucano (Aikhenvald 2003; Zúñiga 2007). Aikhenvald (2003: 142) states that Tariana uses different case-marking based upon the topical status of non-subject constituents. Two object markers are found. The first is *-na*, which is used with pronominal elements along with a cross-referencing prefix. This marker covers a variety of relations, including direct objects, recipients and themes in ditransitive constructions, directionals, and possessors standing in a part-whole relationship with the possessed entity (Aikhenvald 1994: 204-206; Aikhenvald 2003: 142), as in (6.78a):

(6.78) Tariana (Arawakan)

*mhaida pi-ka nu-na*  
 PROHIB 2SG-look 1SG-DOM<sub>1</sub>  
 "Don't look at me!" (Aikhenvald 1994: 204)

The other marker, *-nuku* (and its variant *-naku* used by elder speakers), is used with non-pronominal non-subject NPs<sup>16</sup>. In this case, too, the marker is used with direct objects, recipients or themes in ditransitive constructions, provided that these are referential (6.78a vs. b) and/or topical, as in (6.79c), where "the gold" is the topic at the centre of discussion. Indeed, the marker is also compulsory when the NP constitutes the topic of subsequent discourse (see Aikhenvald 1994: 206 ff.):

(6.79) Tariana (Arawakan)

(a) *Panisi nu-ni-naka*  
 house 1SG-do-PRS.VIS  
 "I make houses (in general)"

<sup>16</sup>It originates from a former locative, still present in the related language Baniwa, where *-naku* is a locative case marker meaning "on/to the surface" (Zúñiga 2007).



- (b) *Panisi-nuka nu-ni-naka*  
house-DOM<sub>2</sub> 1SG-do-PRS.VIS  
"I make a house"
- (c) *di-hẽ-ta-pidana* *diha pa:ku-nuku*  
3SG.NF-see-CAUS1-CAUS2-REM.PST.REP ART gold-DOM<sub>2</sub>  
"He (the Makú gold miner) showed the gold (to the white master  
about to kill him)" (Aikhenvald 2003: 145)

Furthermore, *-nuku* may appear on temporal adverbials and directional NPs. The Tariana topical non-subject marker matches exactly the functions of Tucano *-re*. Just as in Tucano, it is obligatory for pronouns and proper names, whereas it can be used with other non-subject constituents only if they are topical and definite. It is also interesting to note that Tariana exhibits different case-marking on subjects/agents depending upon the contrastive/focal status (Aikhenvald 2003: 140): non-focal subjects are uncoded, whereas focal or relevant newly introduced subjects/agents are overtly coded with the suffix *-nhe*, *-ne*, as in Desano or Tucano. The focal subject marker is also used to disambiguate "who did what as a marker of turn taking" (Aikhenvald 2003: 140) and never co-occurs with the topical non-subject marker *-nuku*. The distribution of case-marking in Tariana is given in Table (6.9):

Topical Non-Subjects	Focal S/A
<i>-na</i> (pronominal) <i>-nuku</i> (non pronominal)	<i>-nhe</i> , <i>-ne</i>

Table 6.5: Distribution of case marking in Tariana

Tucanoan languages and Tariana are somehow extreme in their grammaticalisation of information-structure distinctions into syntax. Indeed, these languages have developed and grammaticalised a case system to code explicitly grammatical functions when their discourse properties deviate from the usual distribution: subject/agent: primary topic, direct object: focus/(secondary) topic. In other words, as topicality is a typical feature of subjects/agents, when they are focused, they are overtly coded. By contrast, as direct objects are associated with either focus or secondary topic relations (i.e. they can have medium topicality, cf. Croft 1991: 155), they are overtly coded if they are highly topical. Therefore, they provide further evidence for the key role of the notions of topic and focus in the rise and development of case-marking systems.

Let us now turn to the Arauan language Paumarí.<sup>17</sup> Paumarí (Chapman and Derbyshire 1991, Chapman 2008 [1979]) shows a different alignment pattern based on word order, with an ergative pattern when the basic SVO word order is followed (6.80a). With OSV and SOV word orders, an accusative system is at work, and the direct object can be overtly coded by the enclitic *-ra* (Chapman and Derbyshire 1991: 164 ff.). In this case, the subject/agent can never occur pre-verbally. As Chapman (2008 [1979]) points out, the enclitic *-ra* is used when the direct object is in pre-verbal position and constitutes the (primary) topic, as shown by the examples in (6.80)<sup>18</sup>. Interestingly, DOM is obligatory with object pronouns, which must occur in pre-verbal position, as shown by (6.80b). Therefore, when an overt object pronoun is present, only an accusative system is possible:

## (6.80) Paumarí (Arauan)

- (a) *Dono-a bi-ko'diraha-'a-ha ada isai hoariha*  
 Dono-ERG 3SG-pinch-ASP-THEME.M DEM.M child other  
 "Dono pinched the other boy" (Chapman and Derbyshire 1991: 164)
- (b) *mafo ho-ra anana-ha-'i-hi*  
 ant.PL 1SG-DOM bite-DUR-ASP-THEME  
 "Ants were biting me" (Chapman and Derbyshire 1991: 251)
- (c) *kahami-ra a-na-joi-vini*  
 palm\_nuts-DOM 1PL-CAUS-return-DEP.TR  
 "We returned to get nuts" (Chapman and Derbyshire 1991: 250)
- (d) *'ada baroro 'dara'darahana-ra ni-va-karaga-'i-ki*  
 DEM leaves big/black-DOM NEG-3PL-find-ASP-INDEP  
 "The big black leaves, they didn't find (them)" (Chapman 2008 [1979]: 11)

In her overview of the Paumarí system, Chapman (2008 [1979]: 16) summarises the link between case-marking and information structure in transitive

<sup>17</sup>Due to lack of data, I will not deal with the situation found in Jamamadí here. Jamamadí has an enclitic *-ra* that is used to overtly code direct objects (Campbell B. 1973, Campbell R. 1977). However, as already noted by Derbyshire (1986: 496), it is difficult to determine the exact conditions that determine the presence of the object marker, as neither of the authors of the descriptive studies on Jamamadí delved deeper into them. However, it seems that information-structure parameters may be relevant for the appearance of the marker, which is clearly optional in some contexts. The accusative marker in Jarawara (Dixon 2004) is no longer used by young speakers, but the conditions governing its use seem to overlap somehow with those found in Paumarí with regard to the topic-marking function of *-ra* (Vogel 1989: 45-46).

<sup>18</sup>The enclitic *-ra* encodes recipients and benefactives as well. Another context in which the enclitic may occur is when a non-subject NP is right-dislocated. See Chapman and Derbyshire 1991: 191).

clauses as follows: when the SVO order is followed, the agent is the topic of the clause. When the direct object is found in pre-verbal position (i.e. OSV, SOV), the direct object becomes the primary topic of the clause and the agent is still informationally prominent (i.e. identifiable and given) but does not constitute the topic of the clause, as the centre of attention is shifted to the direct object.

Topicality, besides animacy, is the most important parameter for DOM to appear in Paraguayan Guaraní, as demonstrated by Shain and Tonhauser (2010). Based on a quantitative study of a corpus of naturally occurring data, Shain and Tonhauser show, *contra* Bossong (1985), that definiteness does not affect DOM, as opposed to animacy and topicality (the latter gauged using Givón's (1983) topicality measures of referential distance and topic persistence). As for animate/human direct objects, Shain and Tonhauser's findings confirm the idea, already found in Gregorez and Suárez's (1967) grammar, that DOM in Guaraní is closely linked to animacy. To be more precise, DOM is much more frequent with human direct objects (6.81a, vs. b and c, see below), even though overtly coded non-human and inanimate direct objects may be found (6.81d, e):

## (6.81) Guaraní (Tupi, Tupi Guaraní)

- (a) *o-hecha* *Juán-chi* *ha* *Pirúlo-pe* *o-ñe-moĩ* *o-hupi*  
 A3-see *Juán*-DIM and *Pirulo*-DOM A3-REFL.PASS-put A3-raise  
*i-po* *ichupe*  
 B3-hand PRON.OBJ.3  
 "It saw Juanito and Pirulo getting ready to say goodbye"
- (b) *o-hecha* *ju'i* *o-po-po* *o-hó-vo* *h-apykuéri-kuéra*  
 A3-see frog A3-jump-jump A3-go-when B3-behind-PL  
 "He saw the frog coming, jumping and jumping behind them"
- (c) *o-mo-ngakuaa* *karai* *pe* *mitã*  
 A3-CAUS-grow gentleman that child  
 "The gentleman raised that child"
- (d) *Juán-chi* *nd-o-jurá-i* *ju'í-pe*  
*Juán*-DIM NEG-A3-grab-NEG frog-DOM  
 "Juan did not get the frog"
- (e) *Pe* *tahachi* *n-oi-pysyrõ-i* *kuri* *pe*  
 that police\_officer NEG-A3-save-NEG back\_then that  
*mohenda-há-pe*  
 locate-NMLZ-DOM  
 "That police officer didn't save the computer"

Shain and Tonhauser's study clearly shows, through a statistical analysis of the corpus they collected, that overtly coded direct objects are more topical than the uncoded ones, as they display higher topic persistence in subsequent discourse. In addition, they argue that topicality explains the variation found among human direct objects, as shown by example (6.81c). Indeed, the more topical direct objects are the more likely to be overtly coded.

The last South-American language that will be examined in this section is Mapudungun (Smeets 2008, Zúñiga 2010). Mapudungun shows DOI, encoded by the infix *-fi-* on the verb only with SAP-3 $\leftrightarrow$ 3 configurations i.e. *-fi-* shows up only with third person direct objects<sup>19</sup>. Traditionally, the presence of *-fi-* has been linked to the humanness or definiteness of the direct object (Smeets 2008: 153). In addition, *-fi-* has also an anaphoric function. This contrast is exemplified by (6.82a), where DOI is absent as the direct object is inanimate and indefinite, as opposed to (6.82b), where the direct object is a proper name.

(6.82) Mapudungun (Araucuan)

- (a) *Pe-n*                    *kura*  
 see-1SG.IND stone  
 "I saw a stone"
- (b) *Pe-fi-ñ*                    *Juan*  
 see-OBJ.3-1SG.IND Juan  
 "I saw Juan" (Zúñiga 2010: 143)

However, as pointed out by Zúñiga (2010), these factors are not necessary nor sufficient conditions for DOI to appear, as shown by examples (6.83). In (6.83a), the direct object is indexed on the verb, while in (6.83b) the very same direct object, governed by the same predicate, is not:

(6.83) Mapudungun (Araucuan)

- (a) *Ina-fi*                    *ñi*                    *epu peñi*  
 follow-OBJ.3 POSS.3 two brothers  
 "He followed his two brothers"
- (b) *Feymew ka ina-ka-tu-y*                    *ñi*                    *epu peñi*  
 then other follow-CONT-again-IND POSS.3 two brother  
 "He continued to follow his two brothers" (Zúñiga 2010: 151)

<sup>19</sup>Mapudungun has a direct-inverse system: only direct forms can take *-fi-*, while inverse configurations take the marker *-e-*. See Zúñiga (2006).

Based on the analysis of the appearance of *-fi-* in narrative texts, Zúñiga (2010) demonstrates that the presence of *-fi-* is obligatory only with proper names, thus contravening the prediction that DOM and DOI follow and spread along the animacy hierarchy. The only necessary condition for DOI to appear is that the indexed direct object be a previous topic that will be re-established in the subsequent discourse in a topic-shift construction (Zúñiga 2010: 157), as shown by the passage in (6.84):

(6.84) Mapudungun (Aracauan)

*Kom tripa-rke-i-ngün kintu-a-lu kom püle, welu chew*  
 all go.out-REP-IND-3PL search-FUT-PTCP all POST but where  
*püle rume ka chum-kao no rume*  
 POST even moreover make\_that? NEG even  
*pe-rke-la-fi-ngün tachi pichi küme üllcha domo*  
 see-REP-NEG-OBJ3-3PL ART little good young\_single woman  
*ñam-küle-lu*  
 disappear-RES-PTCP

"All (the relatives) went out to look for (her) everywhere, but wherever they've been, they didn't see the little, beautiful single girl who disappeared" (Zúñiga 2010: 156)

In the previous passage, the girl was the main topic. At the beginning of (6.84), the topic is the family (all the relatives), and then it becomes the girl again, thus giving rise to a topic shift, which is clearly signalled by the presence of DOI. The same situation comes about in the subsequent passage, where the family takes over the role of topic, and when the topic shifts to the girl, *-fi-* appears again. As we have seen above, in many languages DOM and DOI seem to be related to a topic shift/contrastive function. The Mapudungun data thus lend further support to the hypothesis that DOM and DOI are fundamentally driven by information structure, and are subsequently extended through a conventionalisation of DOM and DOI with the typical semantic features of topics.

## 6.10 North and Central America

In this section I will deal with languages from North and Central America, namely Zenzontepec Chatino (Oto-Manguean), Purepecha and Takelma (both isolates), Yakima Sahaptin (Penutian), Choctaw and Koasati (Muskogean), Babine Witsuwit'en and Slave (Na-Dene). Let us start with Zenzontepec Chatino (Carleton and Waksler 2000, 2002). DOM in Zenzontepec Chatino is obligatory only with

pronouns (6.85a), while it is optional with every other kind of direct object. Therefore, as argued by Carleton and Waksler (2002), DOM in Zenzontepec Chatino cannot be accounted for by taking into account solely semantic features like animacy or definiteness, as both animate and inanimate direct objects may or may not be overtly coded, as shown by (6.85, b, c):

(6.85) Zenzontepec Chatino (Oto-Manguaeon, Zapotecan)

- (a) *nka-to-kachiʔqʔ jʔʔi-nqʔ*  
 COMP-hide-3PL DOM-1SG  
 "They hid me" (Carleton and Waksler 2000: 392)
- (b) *nka-to-kachiʔyu nkwitzá*  
 COMP-hide-3SG child  
 "He hid the child" (Carleton and Waksler 2000: 392)
- (c) *Juan ʔ-yuʔu-nto:-yu (jʔʔi) María*  
 Juan COMP-have-eye-3SG DOM María  
 "Juan recognised María" (Carleton and Waksler 2002: 159)

That DOM in Zenzontepec Chatino is primarily governed by topicality is confirmed by the following examples. In (6.86a), the direct object "fruit" has been introduced for the first time in the text and is therefore uncoded, as it is new. By contrast, in (6.86b), the referent of the direct object has been already established and is therefore overtly coded because of its topical, given status:

(6.86) Zenzontepec Chatino (Oto-Manguaeon, Zapotecan)

- (a) *n-te-su nchiʔyu n-te-su nchiʔyu na*  
 HAB-cut.3SG fruit HAB-cut-3SG fruit NEG  
*nt-yotʔi-na tukwi nchiʔyu nte*  
 HAB-know-1PL what fruit HAB.be  
 "He cuts fruit, he cuts fruit; we don't know what kind of fruit it is"
- (b) *nku-tyeina ntu-su-kaʔa na nyate-ʔʔ jʔʔi na*  
 COMP-begin HAB-cut-again ART person-SPEC DOM ART  
*nchiʔyu*  
 fruit  
 "The man began to cut the fruit again" (Carleton and Waksler 2002: 167)

In addition, in Chatino too, the presence of DOM is closely connected to dislocation. Chatino is a VSO language, and initial position is reserved to the most

topical participant in the clause. In Chatino, similarly to what I have observed in many languages, a dislocated direct object must be obligatorily marked, as shown by (6.87a). Interestingly, the preposition can be moved along with the direct object or be stranded in final position:

(6.87) Zenzontepec Chatino (Oto-Manguaeon, Zapotecan)

(a) *jiʔi kiʔyu nka-ra kuŋaʔa*

DOM man COMP-hit woman

"The woman hit the man"

(b) *kiʔyu nka-ra kuŋaʔa jiʔi*

man COMP-hit woman DOM

"The woman hit the man" (Carleton and Waksler 2000: 396)

Purepecha is an isolate language spoken in Mexico (Chamoreau 1999, 2000). At first glance, DOM seems to be governed by humanness, since all human direct objects are always overtly coded (Chamoreau 1999: 100), even when they are indefinite or generic, as exemplified by (6.88a, b):

(6.88) Purepecha (isolate)

(a) *tatakasapifu kuʔatʃa-ʃin-ti tata-mpa-itʃa-ni*

boy obey-HAB-DECL.3SG father-POSS.REL-3PL-DOM

"The boy obeys his parents"

(b) *ife-f-ka-ni ma nanaka-ni*

see-AOR-DECL.1/2SG-1SG a young\_girl-DOM

"I saw a young woman" (Chamoreau 1999)

Both animate and inanimate direct object may be overtly coded, as long as they are definite and/or specific (6.89a, b). Even mass nouns may occur with DOM, as in (6.89c), where the direct object *etferi* "land", which is modified by a demonstrative and a possessive pronoun, is overtly coded. The presence of DOM with mass nouns, which are situated at the very end of the animacy hierarchy, is very uncommon cross-linguistically. A possible explanation could be associated with the information status of the overtly coded direct object. There is indeed a strong influence of topicality on the presence of DOM. When direct objects are dislocated they get overtly coded, as in (6.89d). Example (6.89f) shows the obligatory overt coding for topicalised direct objects, as opposed to (6.89e), where DOM is not required:

(6.89) Purepecha (isolate)

- (a) *ife-fa-ka-ni* *misitu-ni*  
 see-PROGR-DECL.1/2SG-1SG cat-DOM  
 "I'm about to see the cat"
- (b) *pauani-kini* *ints-ku-a-ti-kfi* *inte-ni*  
 tomorrow-2SG.OBJ give-3SG.OBJ-FUT-DECL.3SG-3PL DEM-DOM  
*itfuskuta-ni*  
 tortilla-DOM  
 "Tomorrow, they'll give you this tortilla"
- (c) *p<sup>h</sup>aʔi-f-ka-ni* *ima etferi-ni xutfiti*  
 touch-AOR-DECL.1/2SG-1SG DEM land-DOM 1SG.POSS  
 "I finally set foot on my native soil"
- (d) *uaʔiti-ni* *tata-k<sup>h</sup>eri aʔi-f-ti*  
 woman-DOM old-man tell-AOR-DECL.3SG  
 "The woman, the old man told her" (Chamoreau 1999)
- (e) *Pedru míti-h-ti* *eski Juanu kaká-ka* *má*  
 Pedro know-PFV-3SG.IND that Juan break-3SG.SUBJ one  
*tsúntsu-(ni)*  
 pot-(DOM)  
 "Pedro knows that Juan broke a pot" (Vázquez-Rojas Maldonado 2010)
- (f) *Má tsúntsu-\*(ni) Pedru míti-h-ti* *eski Juanu*  
 one pot-DOM Pedro know-PFV-3SG.IND that Juan  
*kaká-ka*  
 break-3SG.SUBJ  
 "One pot, Pedro knows that Juan broke it" (Vázquez-Rojas Maldonado 2010)

DOM in Purepecha thus exemplifies the process of extension we have seen at work in many languages. DOM grammaticalises with the topic-worthiest direct objects, such as pronouns and human direct objects. Nevertheless, topicality still remains the most relevant factor with the NP classes with which DOM is optional, such as animate and inanimate direct objects. The mandatory presence of DOM when the direct object is topicalised underpins the importance of topicality.

Yakima Sahaptin (Jansen 2010) is particularly interesting. Yakima, like other Sahaptin languages, has obligatory DOM with third human direct objects, as in (6.90a). Animate and inanimate direct objects may be overtly coded (6.90b, c; see Jansen 2010: 322 ff. for more examples):

(6.90) Yakima (Penutian, Sahaptin)



- (a) *i-k'ínu-sha*                      *Maali Sam-nan*  
 3SG.SUBJ-see-IPFV    Mary    Sam-DOM  
 "Mary sees Sam" (Jansen 2010: 135)
- (b) *níkwit-nan tîin-ma pa-tkwáta-xa*  
 meat-DOM    Indian-PL    3PL.SUBJ-eat-HAB  
 "People eat the meat" (Jansen 2010: 139)
- (c) *aw-pam á-pátuk-ta ílkwas-(nan)*  
 now-2PL    3OBJ-place-FUT    wood-DOM  
 "Now set up a pole" (Jansen 2010: 322)

DOM is also used with applicative constructions, as is exemplified in (6.91a). Furthermore, DOM is obligatory with the inverse construction (i.e. when the direct object outranks the agent in animacy and topicality, as in 6.90b, c), while it is not required with the direct construction (i.e. when the agent is higher in animacy and topicality than the direct object; cf. Givón 2001: 154 ff. and Rude 1994: 117 on Sahaptin):

(6.91) Yakima (Penutian, Sahaptin)

- (a) *i-walptáyk-twii-sha*                      *Chuuli-nan*  
 3SG.SUBJ-sing-APPL-IPFV    Julie-DOM  
 "She is singing with Julie" (Jansen 2010: 316)
- (b) *pá-náktkwanin-xa pt'íniks-nan áyat-in*  
 INV-care\_for-HAB    girl-DOM    woman-3>3ERG  
 "The woman takes care of the girl" (Jansen 2010: 138)
- (c) *pá-k'ínu-sha Maali-in Sam-nan*  
 INV-see-IPFV    Mary-3>3ERG    Sam-DOM  
 "Mary sees Sam" (Jansen 2010: 138)

In fact, DOM seems to be fairly common with non-human animate (and, more rarely, inanimate) direct objects as well. Jansen (2010: 136) argues that the presence of DOM is dictated by topicality and animacy (the latter seems to be more important in direct constructions). The obligatory presence of DOM with inverse constructions may be explained by the highly topical nature of the direct object in inverse constructions. Likewise, applied direct objects are usually correlated with high topicality (Peterson 2007: 88 ff.). DOM thus became mandatory with topical direct objects, like those in inverse and applicative constructions. In (6.92), the direct objects must be obligatorily overtly coded, even though they are non-human animate and inanimate respectively:

## (6.92) Yakima (Penutian, Sahaptin)

*anakú pá-tamanwi-ya íchinak tiichám-nan ku pá-tamanwi-ya*  
 when INV-create-PST this.OBJ earth-DOM and INV-create-PST  
*k'úsík'usi-nan*  
 dog-DOM

"When he created this earth, he also created dogs" (Jansen 2010: 352)

Similar constraints are found in other Sahaptin languages, like Nez Perce (see Rude 1986, 1997) and Klamath (where DOM was obligatory with pronouns and demonstratives, see Rude 1988, and Underriner 2002).

The usage of the accusative case in Choctaw (Muskogean; Broadwell 2006) displays an intriguing optionality. Direct objects are optionally overtly coded, as shown by (6.93a, b; see Broadwell 2006: 32):

## (6.93) Choctaw (Muskogean, Western)

(a) *John-a pisa-tok*  
 John-DOM see:N-PST  
 "He saw John"

(b) *Ahi-(ya) honni-tok*  
 potatoes-DOM boil-PST  
 "He boiled the potatoes" (Broadwell 2006: 32)

Although Broadwell (2006: 75) contends that the distribution of the accusative case in Choctaw is not comparable to DOM, as it does not seem to be governed by animacy or definiteness, yet this seems to be a clear case of DOM, once the conditions for the appearance of accusative marker are analysed and compared against a cross-linguistic sample. Indeed, Broadwell (2006: 75) says that the majority of the direct objects that receive accusative coding in Choctaw are objects modified by a determiner (thus definite) as in (6.94a). The only context in which DOM becomes obligatory is when the direct object is topicalised and moved to initial position (Choctaw basic word order is SOV), as shown by the contrast between (6.94b), where the accusative is optional and (6.94c), where the omission of the accusative renders the sentence ungrammatical.

## (6.94) Choctaw (Muskogean, Western)

(a) *Makaatokoosh ihasibis-ma takaachi-took miyah*  
 and.then tail-DEM.ACC hang-DIST.PST HSAY  
 "And then he hung it on his tail, they say"

- (b) *John-at takkon-(a) chopah*  
 John-NOM peach-DOM buy-TNS  
 "John bought a peach" (Broadwell 2006: 32)
- (c) *Takkon-\*(a) John-at chopah*  
 peach-DOM John-NOM buy-TNS  
 "John bought a peach" (Broadwell 2006: 39)

As we have seen, DOM is often found with dislocated direct objects, because of their topicality. Indeed, Broadwell himself acknowledges that speakers "interpret NPs with overt accusative marking as topical" (Broadwell 2006: 74) and as the current centre of attention. The Choctaw pattern cannot be explained by appealing to ambiguity avoidance, as subjects/agents are always obligatorily overtly coded<sup>20</sup>. As we have seen in this chapter, the restriction of DOM to dislocated elements is by no means rare across the languages of the world. Furthermore, as I will show in chapter 8 on Romance languages, dislocated direct objects seem to be the first kind of direct objects that receive DOM diachronically.

Now, I will analyse the behaviour of DOI limited to third person in two Na-Dene languages, Slave and Babine-Witsuwit'en (Athabaskan). Athabaskan languages do not have case-marking and use verbal prefixes to index core arguments and some adjuncts. The most common word order is SOV, but other permutations are possible for pragmatic reasons (Rice 1989, Gunlogson 2001). In many Northern Athabaskan languages, such as the ones I am discussing here, third person direct object indexation is in complementary distribution with overt NPs (Rice 2003, Saxon 1989)<sup>21</sup>. That is, when a NP is present, the prefix is obligatorily absent (this complementarity, in fact, does not hold for Babine, see below).

Babine-Witsuwit'en (Gunlogson 2001) shows a more complicated pattern, insofar as third person DOI (coded by the prefixes *y-* and *hiy-*) sometimes can co-occur with overt NPs, as shown by (6.95a) as opposed to (6.95b):

(6.95) Babine Witsuwit'en (Na-Dene, Athabaskan)

- (a) *we-yi-stne'*  
 NEG-3SG.OBJ-3SG.drink  
 "S/he isn't drinking it" (Gunlogson 2001: 385)

<sup>20</sup>There seems to be a similar pattern in Koasati, where the accusative is often omitted when the direct object is adjacent to the predicate (Kimball 1991: 394). However, the available evidence does not allow to draw any conclusion for this language.

<sup>21</sup>Rice (2003) assumes that the complementary distribution between overt NP and indexation markers reflects an old stage among Athabaskan languages, as in Southern Athabaskan languages, e.g. Navajo, subject and object indexation is obligatory. Therefore, the Northern Athabaskan pattern should be the conservative one.

- (b) *Hida dini yi-nilh'en*  
 moose man 3SG.OBJ-3SG.look  
 "The moose is looking at the man" (Gunlogson 2001: 374)

As Gunlogson (2001) points out, the presence of the direct object prefix in examples (6.95a, b) correlates with a definite interpretation. This does not come as a surprise when DOI appears without an overt direct object, since in this case the object prefix functions anaphorically. The strongest evidence for this claim, however, comes from the obligatory occurrence of DOI with proper names (6.96a), demonstrative NPs (6.96b), and possessed direct objects (6.96c), as exemplified in (6.96):

(6.96) Babine Witsuwit'en (Na-Dene, Athabaskan)

- (a) *Ggi lhc Alfred yi-lhc'ilh - \*silhc'ilh*  
 that dog Alfred 3SG.OBJ - 3SG.PST-bite  
 "That dog bit Alfred"
- (b) *Sean ggi c'ilhtiy yi-ka'ninzi/ha'ninzi.*  
 Sean that gun 3SG.OBJ-3SG.PRS-want  
 "Sean wants that gun/a gun (without DOI)"
- (c) *Lillian George bi-ka' we-yu-taskitl*  
 Lillian George 3SG-car NEG-3SG.OBJ-3SG.buy  
 "Lillian isn't going to buy George's car"

Gunlogson (2001: 393) proposes that third person DOI in Babine Witsuwit'en –especially with regard to the co-occurrence of an overt direct object and DOI– serves to overtly code the topicality of the indexed direct objects. This proposal is corroborated by evidence from different domains. First, as Gunlogson (2001: 385–389) argues, indexed direct objects fall outside the scope of negation. The "car" in example (6.96c) is unaffected by the negation, since its existence must be presupposed. As is well known, topical elements must be presupposed, and therefore referents of topical expression are presupposed to exist (Lambrecht 1994: 154).

Second, Gunlogson (2001: 393) demonstrates that DOI in Babine is used when there is a topic shift in the discourse, as shown by (6.97):

(6.97) Babine Witsuwit'en (Na-Dene, Northern Athabaskan)

- Nik dzilh k'it 'awet next day 'awet 'awet ts'intilelh*  
 up mountain on then next day then then 1PL.sleep.PRS  
*tl'i'et'awet didiltlic tl'i'-awet nis kwin nedilhk'aykh*  
 while 3SG.jump\_up.PST and ahead fire 3SG.make.PRS

*tl'i'awet t'alh lha'iwendin'i'lh'iyh. T'alh 'iy-ey'ililh*  
 and-then food 3SG.start\_to\_prepare.PRS food 3SG.OBJmake.3SG.PST  
 "Up on the mountain the next day, while we slept, she'd jump up, make a  
 fire a little way off and then she'd start to prepare food. She made the  
 food" (Gunlogson 2001: 393)

As we can see, when the direct object "food" is first introduced into the discourse, it is not indexed on the verb. In the next sentence, it is instead indexed on the verb via the prefix *iy-*. As observed by Gunlogson, the topic of the story so far has been this woman, *Dilizco*, which is the main topic of the passage in (6.97). The food is introduced at the end of the passage as an indefinite NP. In the next sentence, it is promoted to topic position and therefore indexed on the verb. Indeed, the primary topic of the next sentences becomes the food, as witnessed by the continuation of the story (Gunlogson 2001: 394).

Finally, DOI is found when the direct object is (left)-dislocated, which is a condition for the appearance of DOM and/or DOI that we have encountered in many languages. Indeed, as I have mentioned above, in other Northern Athabaskan languages DOI co-occurs with an overt NP only when the direct object is dislocated, as shown by the following examples from Slave (Rice 1989), which illustrate the presence of DOI when the direct object is dislocated (6.98b, d) and its absence when it is in situ (6.98a, c):

(6.98) Slave (Na-Dene, Northern Athabaskan)

- (a) *Tlį nidq̄q̄ gah tedéhn̄de*  
 rabbit dog far 3SG.PST.chase  
 "The dog chased the rabbit a long way" (no dislocation, no DOI)
- (b) *Gah tlį nidq̄q̄ te-ye-déhn̄de*  
 dog rabbit far 3SG-3SG.OBJ-PST.chase  
 "The rabbit, the dog chased it a long way" (dislocation and DOI)
- (c) *laleníhʔee wehs̄j*  
 sweater 3SG.PST.make  
 "She made a sweater" (no dislocation, no DOI)
- (d) *laleníhʔee-yá ye-hs̄j*  
 sweater-TOP 3SG.OBJ-3SG.PST.make  
 "A sweater she made" (dislocation and DOI; Rice 1989: 1198)

It is noteworthy that the dislocated direct object in (6.98d) bears the overt topic marker. The presence of the topic marker indeed corroborates our hypothesis that it is the topical status of the direct object that brings about the use of DOI. Besides

having DOI with dislocated direct objects, Slave displays another interesting DOI pattern. Although nominal direct objects and DOI are said to be mutually exclusive in Slave, nominal human plural direct objects are indexed on the verb, as shown by (6.99):

(6.99) Slave (Na-Dene, Northern Athabaskan)

- (a) *deneke go-gháyeda/\*gháyeda*  
 people.PL 3PL.OBJ-3SG.see  
 "S/he sees the people" (Rice 1989: 1016)
- (b) *?eyi dene ye-yaake teh-go-déhnde*  
 the man POSS-child.PL 3SG-3PL.OBJ-IPFV.chase  
 "The man chased her cubs" (Rice 1989: 1020)

The data from Babine and Slave are particularly interesting because they allow us to see the different stages of the development of DOI. While Slave appears to be more conservative than Babine, as it maintains DOI only with dislocated, topicalised, direct objects, Babine Witsuwit'en has extended DOI from topicalised direct objects to topic-worthy direct objects, such as proper names and demonstratives, regardless of their actual topicality. However, it turns out that DOI in Babine still carries out a topic-shift function, which probably derives from the original dislocated environment in which DOI arose. An additional change is probably taking place in Slave, where human plural direct objects trigger DOI, as opposed to non-human ones.

The last language that I examine in this section is Takelma (isolate, possibly Penutian). My discussion is based on Culy (2000), who based himself on the grammar and the texts collected by Sapir (1922). In Takelma, overt indexation is required for first and second person objects, with singular and plural distinguished. For third person direct objects indexes, Takelma makes use of *-k<sup>h</sup>wa*, which alternates with zero. According to Culy (2000), third person direct objects can be used when the direct object is higher in animacy than the subject, as in (6.100a) vs. (6.100b). Second, if DOI is present, no overt nominal direct object can be used, as in Slave above (this idea is clearly stated by Sapir 1922: 168)<sup>22</sup>.

(6.100) Takelma (isolate)

- (a) *t'ipisi: t'ayák<sup>h</sup>*  
 ant.PL find.PST.3SG?  
 "He found the ants" (Culy 2000: 3)

<sup>22</sup>In fact, it seems that there are instances in the text analysed by Culy where direct objects co-occur with DOI. However, there is not enough evidence to draw any conclusions from those examples, which Culy considers spurious

- (b) *t'ipisi: t'ayá-k<sup>h</sup>wa*  
 ant.PL find.PST.3SG?-3.OBJ  
 "The ants found him" (Culy 2000: 3)

Third, in Culy's analysis, DOI is used when the direct object is topical. DOI would be triggered because the direct object has already been mentioned and is therefore at the centre of attention, i.e. topical. This is why the direct object in (6.101) is indexed on the verb:

(6.101) Takelma (isolate)

- há:xan-k<sup>h</sup>wa-hi:s*  
 burn-3.OBJ-almost  
 "He (Sinew-man) almost burnt him (Daldal)" (Culy 2000: 5)

Most probably, based on what we have seen in other languages, one could contend that Takelma DOI was indeed conditioned by topicality. However, the available data make it impossible to provide a deeper characterisation of the distribution of DOI in Takelma.

## 6.11 Austro-Asiatic, Austronesian and Papuan

In this section I will consider languages from the Pacific Ocean area, plus the Austro-Asiatic language Semelai. Except for Manambu (Sepik, Ndu) and Namia (Sepik, Yellow River), all the other languages that will be examined in this section are part of the Austronesian family (Tawala, Tamabo, Tinrin, Nêlêmwa, Selayarese, Malagasy, Tobati, Marquesan).

Let us start with Semelai. Semelai codes post-verbal direct objects with the proclitic *hn-*, as shown by (6.102):

(6.102) Semelai (Austro-Asiatic, Mon-Khmer)

- (a) *ki-tikam hn-kɔʔ ke*  
 3AG-stab DOM-pig\_tailed\_macaque that  
 "He stabbed that pig-tailed macaque" (Kruspe 1999: 261)
- (b) *ki-yɔk la-kniək hn-bantal*  
 3A-take A-husband DOM-pillow  
 "The husband fetched a pillow" (Kruspe 1999: 91)
- (c) *d-ɔs hɛʔ dɔl ki-pan-cin-hn*  
 reach LOC.above house 3AG-CAUS-cook-DOM  
 "(When he) arrived home, he cooked (it)" (Kruspe 1999: 91)

Kruspe (1999: 262) states the DOM in Semelai is optional in all contexts, as there seems to be no influence of factors like animacy or affectedness. Crucially, the only context in which DOM is disallowed is when the direct object is pre-verbal (this applies to the agent as well, since all pre-verbal constituents lose their case-marking). In addition, overtly coded direct objects were never elicited, but occurred only in spontaneous discourse. Another interesting context in which DOM is present is exemplified by (6.102c) above, where *-hn* is encliticised to the verb and refers anaphorically to a non-overtly expressed direct object. It may well be, thus, that the distribution of DOM in Semelai is still very pragmatic in nature, as the anaphoric function we have just seen suggests.

The clitic *hn-* has a wide array of functions. Besides optionally coding direct objects, the very same clitic is used to code the third person possessor and pronominal indirect objects, as well as on conditional and temporal conjunctions (Kruspe 1999: 382). These usages could be connected to topicality, as demonstrated by the frequency of this polysemy cross-linguistically<sup>23</sup>.

Manambu has both DOM and DOI (Aikhenvald 2008). Both of them are conditioned by the topicality of the direct object referent. DOM is coded by the accusative suffix *-Vm* which is used for the locative as well. As a general rule, DOM occurs with definite and specific direct objects. However neither definiteness nor specificity alone can predict the appearance of the marker, since pronouns, which are usually overtly coded in Manambu, can show up unmarked under exceptional circumstances (see below). This is exemplified by (6.103a), where the pronominal direct object is uncoded, as opposed to (6.103b). Example (6.103c) shows both DOM and DOI with a definite human direct object:

(6.103) Manambu (Sepik, Ndu)

- (a) *wun ñən ma və:*  
 1SG 2SG NEG see.NEG  
 "I didn't see you" (Aikhenvald 2008: 145)
- (b) *wun-a:m mən karda ma: ta:y*  
 1SG-LK:DOM 2SG:M bring.down NEG first  
 "You will not carry me down first"
- (c) *dakul wapi du-a-ñanugw-a:m*  
 spirit bird man-LK-children-LK:DOM  
*kə-da:-di*  
 eat-3PL.SUBJ.PST-3PL.OBJ.PST

<sup>23</sup>There are further usages that are rather puzzling. Indeed, *hn-* is used in "agent suppressed" constructions such as middle voice constructions. It would thus seem that, when the agent is suppressed, the other participant retains its coding irrespective of the change in verbal voice (Kruspe 1999: 91).



"The spirit birds ate up male children" (Aikhenvald 2008: 149)

Interestingly, the direct object can be indexed on the verb when it is more topical than the agent. However, the Manambu indexation system is more complicated, since it follows a subject/non-subject principle. While the subject/agent must be obligatorily indexed on the verb, other non-subject constituents (including direct and indirect objects, obliques such as time, location, destination and manner constituents) are indexed on the verb as long as they are topical (Aikhenvald 2008: 61-62). In addition, DOI becomes obligatory when the direct object is accompanied by a "reactivated topic" demonstrative, i.e. a topic marker employed to re-introduce a previously mentioned topic that has been absent for some time in the discourse (Aikhenvald 2008: 69 ff; 219 ff.)<sup>24</sup>.

As for DOM, Aikhenvald (2008) clearly shows that it is governed by the topicality of the direct object: since topical elements must be definite and specific, the apparent connection with these factors is easily accounted for. The importance of topicality in determining the overt coding of a direct object is confirmed by the absence of DOM on focal direct objects. Indeed, DOM is never found when the direct object is in a contrastive focus construction (which Aikhenvald (2008: 540) calls "highlighting focus"), as shown by example (6.104a, b):

(6.104) Manambu (Sepik, Ndu)

- (a) *[du də-kə-də kui-ad]*<sub>FOC</sub>  
 man 3SG-OBL-M.SG meat-3SG.M.NOM  
*kə-da-d*  
 eat-3PL.SUBJ.PST-3SG.M.BAS.PST  
 "It is man's flesh they ate" (Aikhenvald 2008: 541)
- (b) *də-kə-di ñan-ugw kur-taka wula-taka [Apur]*<sub>FOC</sub>  
 3SG-OBL-PL child-PL get-IMM.SEQ enter-IMM.SEQ Apur  
*Iraman vya-də-d*  
 Iraman hit-3SG.M.SUBJ.PST-3SG.M.BAS.P  
 "Having taken his children, having entered (the battlefield), Irama hit Apur (not anyone else)..." (Aikhenvald 2008: 536)

Summing up, we have seen that the presence of DOM and DOI in Manambu strongly correlates with topicality. Likewise, topicality is the trigger for DOM

<sup>24</sup>Interestingly, these demonstratives/topic markers can only refer to subjects and direct objects. This fact confirms the strong affinity between these grammatical relations and the (re)-introduction of new or previously mentioned participants in discourse observed by Dubois (1987) even in a nominative-accusative language like Manambu.

in Namia too (Tupper 2009). DOM in Namia is compulsory with pronouns<sup>25</sup>, proper names (6.105a) and human nouns. It can also be found on every kind of NP, including abstract nouns, provided that they are referential, as shown by the opposition between (6.105b) vs, (6.105c):

(6.105) Namia (Sepik, Yellow River)

- (a) *Aya, er Amae-mə ta-na-laol-wa*  
 father 1DU Amae-DOM SPEC-REFL-trade-IRR  
 "Father, may we two exchange Amae?"
- (b) *mokuran almar-mə pə-na-we*  
 REF.PL poison\_vine-DOM SEQ-dig.out-REAL  
 "Someone dug out poison vines"
- (c) *almar pə-na-we lommomə*  
 poison\_vine SEQ-dig.out-REAL 3PL.DOM  
 "The ones who were digging out poison vines" (Tupper 2009)

Topicality seems to have an important role in the overt coding of direct objects, since DOM is found with every kind of direct object, as long as it is construed as topical. Moreover, when the direct object is topicalised and the standard SOV order is not followed, DOM is required, as in (6.106):

(6.106) Namia (Sepik, Yellow River)

- (a) *[Balira na-mə]<sub>TOP</sub> mi liranía pə-nake-irl-e*  
 plane also-DOM fire all SEQ-ACCOM-burn-REAL  
 "Fire burnt up the plane as well (lit. The plane too, the fire burnt)"
- (b) *[Tapo-la-mə]<sub>TOP</sub> Witjowe-la-kə ar Iwae Temau maem*  
 Tapo-M-DOM Witjowe-M-OBL dog pig Temau at  
*aro i-re*  
 TEMP.DEM bite-REAL  
 "The pig (found by) Witjowe's dogs bit Tapo at Temau" (Tupper 2009)

As Tupper points out, the leftmost position in the clause is reserved to topics and constitutes the sole means to promote to topic a constituent other than the subject/agent. The strong tendency to overtly code topical direct objects in topic-promoting constructions seems to be confirmed for Namia as well. As we will

<sup>25</sup>Third person pronouns have fused with the object marker to produce distinct object pronouns, as in (6.105c).

see shortly, this correlation holds for the Austronesian languages surveyed in this section.

Tinrin (Austronesian, Malayo-Polynesian) is an exemplar case in this respect. Tinrin does not have either DOM or proper DOI (Osumi 1995). Its standard word order seems to be VOS (Osumi 1995: 223). When an animate direct object is topicalised, however, there must be a resumptive pronoun after the verb (6.107b, c), as opposed to inanimate direct objects (6.107d), which cannot co-occur with the resumptive pronoun, as shown by the examples in (6.107):

(6.107) Tinrin (Austronesian, Malayo-Polynesian)

- (a) *haru ru jorri tau nrî ârijù*  
 1DU 1DU see often 3SG down.there  
 "We two often saw him down there" (VOS order; Osumi 1995: 206)
- (b) *Sonya, nrâ ta nrî nrâ*  
 Sonya, 3SG hit 3SG PST  
 "Sonya, he hit her" (OSV order, animate direct object; Osumi 1995: 241)
- (c) *bwò ri ta rri ru meemarri*  
 crab 1PL.INCL catch 3PL at reef  
 "Crabs, we catch them at the reef" (OSV order, animate direct object; Osumi 1995: 241)
- (d) *wa mi ha nrâ jurrù nrâ ausòò-rò*  
 ART watermelon PROX 3SG cut SUBJ elder\_sibling-1SG  
 "This watermelon, my elder brother cut (it)" (OSV order, inanimate direct object; Osumi 1995: 206)

A similar pattern is found in Nêlêmwa (Bril 2004). As Osumi and Bril argue, these mechanisms are used to promote an entity to topic position. Tinrin and Nêlêmwa appear to be developing an incipient DOI system which, unsurprisingly, is starting when the direct object is promoted to topic position via left dislocation. Similarly, in Tamabo (Oceanic), there is complementary distribution between object suffixes and overt direct objects. However, DOI becomes compulsory when the direct object is dislocated (Jauncey 2002: 622).

Now I will turn to two Austronesian languages which show DOI solely with definite direct objects, i.e. Selayarese (Sulawesi) and Tawala (Eastern Polynesian). Selayarese is held to be a morphologically ergative language, which indexes agents via a prefix, and subjects/direct objects via a suffix (Finer 1997). When the direct object is definite, there is DOI on the verb (6.108a, b). If the direct object is indefinite, the verb takes the intransitive suffix (6.108c):

## (6.108) Selayarese (Austronesian, Sulawesi)

- (a) *la-ʔalle-i*                      *doeʔ-ñjo*    *i*        *Basoʔ*  
 3SG.SUBJ-take-3SG.OBJ money-the HUM Baso'  
 "Baso took the money"
- (b) *la-keoʔ-a*                      *i*        *Basoʔ*  
 3SG.SUBJ-call-1SG.OBJ HUM Baso'  
 "Baso called me"
- (c) *(a)ng-alle-i*    *doeʔ*    *i*        *Basoʔ*  
 INTR-take-3SG money HUM Baso'  
 "Baso took (some) money" (Finer 1997: 679-680)

So far, it would seem logical to assume that DOI in Selayarese is determined by the formal definiteness of the direct object. In fact, the situation is a little more complex and the governing property seems to be, once again, the topical status of the direct object. I will briefly review the evidence that supports this claim. First, DOI is prohibited with the *wh*-word *apa* "what", as shown by (6.109); Basri and Finer (1987: 144).

## (6.109) Selayarese (Austronesian, Sulawesi)

- apa la-taro-(\*)i*                      *ri lamari*    *i*        *Basoʔ*  
 what 3SG.SUBJ-put-3SG.OBJ in cupboard HUM Baso  
 "What did Baso put in a cupboard?" (Basri and Finer 1987: 144)

Furthermore, as discussed by Basri and Finer (1987: 145), when an indefinite (specific) direct object is moved to initial position (i.e. it is topicalised), the verb cannot take the intransitivising prefix and the normal transitive construction must be used. That indexation is absent with focal constituents is not so surprising, given that all the factors favouring DOI are usually closely associated with topicality rather than with focality (see Siewierska 2004: 159-163 for a thorough discussion and examples).

Tawala (Eastern Polynesian; Ezard 1991) has a DOI system that encodes number and person of animate and definite direct objects, as exemplified in (6.110). Examples (6.110c, d) show the absence and the presence of DOI with an indefinite non-specific and a definite direct object respectively:

## (6.110) Tawala (Austronesian, Eastern Polynesian)

- (a) *Ega tsu o-ne-bahe-bah'-e-u*  
 NEG 1SG 2PL.SUBJ-POT-DUR-say-TRZ-1SG.OBJ  
 "Don't talk about me"

- (b) *Meyagai lawa-hi i-pali-ye-hi*  
 village person-3PL 3SG.SUBJ-scold-TRZ-3PL.OBJ  
 "She rebuked the village people" (Ezard 1991: 100)
- (c) *Ta-nae polo ta-lugowada*  
 1PL.INCL.SUBJ-go pig 1PL.INCL.SUBJ-steal  
 "Let's go pig stealing"
- (d) *Ta-nae Kama a polo*  
 1PL.INCL.SUBJ-go Kama his pig  
*ta-lugowad'-i*  
 1PL.INCL.SUBJ-steal-3SG.OBJ  
 "Let's go and steal Kama's pig" (Ezard 1991: 96)

Ezard (1991: 93-95) observes that the presence of DOI on direct objects correlates with topicality. As a matter of fact, DOI is present only when the direct object is topical, as in (6.111a), while it is absent when the direct object is focal (6.111b). Furthermore, he notes that DOI is consistently used when the direct object is topicalised (Tawala word order being SOV), as in (6.111c)<sup>26</sup>:

(6.111) Tawala (Austronesian, Eastern Polynesian)

- (a) *Wam a-gelu-ya*  
 boat 1SG.SUBJ-embark-3SG.OBJ  
 "I embarked the boat"
- (b) *Wam amaka hi-gelu*  
 boat already 3PL.SUBJ-embark  
 "They had already boarded a boat" (Ezard 1991: 95)
- (c) *Polo Kukuku ega i-ta-uni-i*  
 pig Pheasant NEG 3SG.SUBJ-IRR-catch-3SG.OBJ  
 "The pig wasn't caught by Pheasant" (Ezard 1991: 94)

In contrast to the Austronesian languages I have analysed thus far, Marquesan and Malagasy display DOM. Let us first discuss Malagasy. As I have already mentioned, Malagasy DOM system is based on formal definiteness, rather

<sup>26</sup>It must be noted that Tawala does not possess a passive construction, even though the English translation of (6.111c) is passive. However, the translation itself highlights the fact that the direct object is the primary topic of the sentence. Moreover, Ezard observes that, although the direct object is in initial position, the intonation pattern remains the same as in the standard SOV word order, i.e. the direct object in these constructions is not set off from the rest of the sentence by a break.

than information-structural notions like identifiability or topicality. DOM, expressed via the preposition *an*, is used with proper names (6.112a), (optionally) with demonstratives (6.112b) and with NPs containing the article *ilay* (6.112c)<sup>27</sup>, which is employed when the NPs refers back to a recently mentioned entity (Keenan 2008: 245; Zribi-Hertz and Mbolatianavalona 1999; Pearson 2001). In addition, Malagasy object pronouns have an initial *a* or *an* that can be traced back to the preposition coding nominal direct objects (6.112d). As pronouns are in the highest position in the animacy/definiteness/topicality hierarchies, it is not unexpected that they incorporated the preposition into the paradigm (Zribi-Hertz and Mbolatianavalona 1999: 193), making it an obligatory element with pronouns:

## (6.112) Malagasy (Austronesian, Barito)

- (a) *Nanenjika an-dRabe aho*  
 PST.AV.chase DOM-Rabe 1SG  
 "I chased Rabe"
- (b) *Tsy mahalala (an) io olona io aho*  
 NEG PRS.AV.know DOM that person that I  
 "I don't know that person"
- (c) *Nisy nandray an ilay sary teto*  
 was PST.AV.take DOM that picture PST.here  
 "Someone took that picture that was here"
- (d) *Nanenjika ahy izy*  
 PST.AV.chase 1SG.ACC 3SG  
 "He chased me" (Keenan 2008: 245)

DOM is prohibited in two contexts. The first is when the direct object is preceded by the definite article *ny*<sup>28</sup>, as in (6.113a), and when the direct object is bare, as in (6.113b):

## (6.113) Malagasy (Austronesian, Barito)

- (a) *Tsy mahalala (\*an-) ny anadahin'i Soa aho*  
 NEG know DOM the brother.OF.ART Soa 1SG  
 "I don't know the brother of Soa"

<sup>27</sup>While Keenan (2008: 245) contends that DOM with the previous mention article is optional, Zribi-Hertz and Mbolatianavalona (1999: 191) maintain that DOM must be used when the direct object phrase contains it.

<sup>28</sup>Zribi-Hertz and Mbolatianavalona (1999: 186) note that the definite determiner *ny* is disallowed in direct object position altogether, unless the direct object is not modified by a possessive NP, as in (6.113b).

- (b) *Manao* (\*an-) *farafara mahafinaritra io mpandrafitra*  
 PRS.AV.make DOM bed pleasing that carpenter  
*io*  
 that

"That carpenter makes pleasing beds" (Keenan 2008: 246)

The complementary distribution of the definite article and DOM has already been observed in Chapter 5 with regard to Zuni and Corsican (Romance), and a tentative diachronic analysis of this rather unusual pattern has been proposed there.

As far as DOM in Malagasy is concerned, it seems that the tendency is not too dissimilar from the languages surveyed in this chapter. DOM is sensitive to definiteness, and its DOM pattern probably arose through the same discourse conditions we have seen at work in other languages and was later narrowed to definite direct objects only. Indeed, DOM is used with pronouns (with which it has been completely grammaticalised), with proper names and with NPs that have been already introduced into the discourse. However, more investigations are needed to work out the exact conditions for the appearance of DOM in Malagasy.

The last Austronesian languages I examine here are Marquesan (Oceanic; Cablitz 2006), and Tobati (Donohue 2002). Marquesan overtly codes pronouns, proper names, and definite direct objects via the preposition *i* (*ia* before pronouns and proper names), which is also used for indirect objects<sup>29</sup>, as shown by the following examples:

(6.114) Marquesan (Austronesian, Oceanic)

- (a) *Ua to'o naiho ia ia ua ta, u kukumi, ua kai*  
 TAM take just DOM 3SG TAM hit TAM kill TAM eat  
*ia Kanakete*  
 DOM Kanakete

"(They) just took him, (they) just hit him, (they) killed him, (they) ate Kanakete" (Cablitz 2006: 145)

- (b) *I popahi ai Pae'tini i t-a ia tau*  
 TAM command ANAPH Pae'tini DOM ART-POSS 3SG PL  
*manu*  
 bird

"Pa'etini commanded his birds"

<sup>29</sup>The allomorph *ia* probably retains the so-called personal article *a*, now obsolete in Marquesan (Cablitz 2006: 145).

- (c) *To'o Tahia i te-a vehie a Teiki e*  
 take Tahia DOM ART-DEM firewood POSS Teiki TAM  
*koti-koti ei...*  
 RED-cut now

"Tahia takes that firewood of Teiki which (he) has just cut..." (Cablitz 2006: 65)

- (d) *E koti a'a Teiki te vehie me te toki*  
 TAM cut DEM Teiki ART firewood with ART axe

"Teiki is cutting the firewood with an axe" (Cablitz 2006:80)

As examples (6.114c) and (6.114d) show, the very same NP can be overtly coded or uncoded without any difference in verbal semantics or the semantic properties of the direct object referent, since both are clearly definite and specific. I surmise, based on comparative evidence, that these differences in object encoding are brought about by differences in the topicality of the direct object. Let us briefly review the evidence that argues in favour of an influence of topicality on the DOM systems of the type found in Marquesan as well as in other Polynesian languages (see Chung 1978 on relics of this construction in Samoan, Ball 2007 for examples from Rennellese). The preposition employed for DOM is originally a locative preposition, and in many languages including Marquesan, *i* still retains its basic locative/temporal meaning. Interestingly, the preposition *i/ia* has come to be employed as a (contrastive) topic-(shift) marker in other Polynesian languages. For example, Maori uses *ia* as a contrastive topic marker, in addition to its use as a direct/indirect object marker. Moyse-Faurie (2004) documents that this strategy is used also in Wallisian (also known as East Uvean, 6.115) and in West Uvean (6.116). In this latter language, *ia* shows the same restriction I have observed in Marquesan, since it can only occur before pronominal or proper nouns topics (6.116).

(6.115) Wallisian (Austronesian, Oceanic)

- ia tana kui 'e manatu ki ai ia Soana*  
 TOP POSS grand\_parent INACC think OBL ANAPH ABS Soana  
 "As for his grandparents, Soana thinks of them" (Moyse-Faurie 2004)

(6.116) West Uvean (Austronesian, Oceanic)

- Ga ia Ludovic naia too-tama-ina a Pascal*  
 and TOP Ludovic passed 3SG.take-child-TR ART Pascal  
 "As for Ludovic, he has adopted Pascal" (Moyse-Faurie 2004)



As pointed out by Claire Moyse-Faurie (personal communication), since the allomorph of the locative marker *i* before pronouns and proper nouns is *ia*, it is indeed very likely that the topic marker *ia* comes from the locative marker. In Puyuma (Austronesian, Puyuma), *i* serves to encode fronted topics, as in (6.117b), where the NP is fronted, as opposed to (6.117a):

(6.117) Puyuma (Austronesian, Puyuma)

- (a) *ta-ilrang-aw dra enay na driketr-an*  
 1PL.GEN-grind-TRANS1 ID.OBL water DEF.NOM sticky-NMLRZ  
 "We grind the sticky rice with water"
- (b) *na drikedr-an i ta-ilrang-aw dra*  
 DEF.NOM sticky-NMLRZ TOP 1PL.GEN-grind-TRANS1 ID.OBL  
*enay*  
 water  
 "The sticky rice, we grind it with water" (Teng 2008: 151-152)

This development, as well as the comparative evidence I have adduced in this chapter, provides strong support for the hypothesis that DOM in Marquesan relies upon the topicality of the direct object.

The last language I discuss in this section is Tobati, where DOM is found when the standard OSV word order is not followed (Donohue 2002: 191):

(6.118) Tobati (Austronesian, Oceanic)

- (a) *Man har-ad rom-ra yar*  
 bird person-DOM see-SEQU fly  
 "The bird saw the man and then flew off"
- (b) *Nehu man rosi-(ad) j-om-rie*  
 1SG bird two-DOM 1SG-see-3PL  
 "I saw two birds" (Donohue 2002: 191-199)

Once again, DOM correlates with the position of the direct object within the sentence. Whilst there is not enough data to detect the general constraints of DOM in Tobati (especially with regard to the differences in information structure related to the variations in word order), it seems plausible, based on the data presented so far, that the appearance of DOM correlates with topical, presupposed direct objects.

## 6.12 Tibeto-Burman

Many languages of East and South-East Asia show DOM patterns based on topicality. In this section I will concentrate my attention on the Tibeto-Burman languages Galo, Dolakha Newari, Burmese, Chepang, Lahu, Chantyal. Although the connection with topicality in some languages is on its way to becoming grammaticalised as an opposition between definiteness/indefiniteness or animacy/inanimacy, still the general relevance of topicality can be ascertained diachronically, as will be discussed below. A good survey of DOM in Tibeto-Burman is provided by LaPolla (1992, 2004), where DOM is referred to as "anti-ergative/agentive marking", and is characterised as a means to disambiguate between semantic roles especially in the case that the direct object is human/animate.

Let us start with Galo (Post 2007). DOM in Galo is expressed by two different forms. The first one, which Post (2007: 721) calls "non-agentive" is exclusively used with animate entities. It obligatorily encodes proper names and given human direct objects (Post 2007: 726), as shown by (6.119a, b):

(6.119) Galo (Sino-Tibetan, Tibeto-Burman)

- (a) *bíi-ká abó-nè gə-dùu*  
 3SG-GEN father-DOM<sub>1</sub> carry/wear-IPFV  
 "He takes after his father"
- (b) *tukáa-nè [...] əm-dùu-nà-əə na*  
 blackie-DOM<sub>1</sub> [...] tell-IPFV-NMZR:SUBJ-COP-IPFV DECL  
 "I've been telling Tuka that..." (Post 2007: 726)

Interestingly, pronouns and other kinds of nominals have a different object marker, realised by the allomorphs *-m* (found with first/second person singular pronouns, reflexives, and simple demonstratives) and *(-)əəm* (Post 2007: 721). The basic factor underlying the distribution of DOM is apparently definiteness. Accordingly, pronouns and demonstratives are invariably overtly coded, as in (6.120a, b), since they are always definite. Other direct objects may be overtly coded as long as they are definite and not-human, as shown by the opposition between (6.119c, d):

(6.120) Galo (Sino-Tibetan, Tibeto-Burman)

- (a) *ŋó nó-m cən-dùu*  
 1SG 2SG-DOM<sub>2</sub> know-IPFV  
 "I know you"

- (b) *higi-m*                      *amó*    *higi-m*                      *na*    *ɲunù-kə̀*  
 SPRX.IND-DOM<sub>2</sub> paddy SPRX.IND-DOM<sub>2</sub> DECL 1PL-GEN  
*lə̀-tà-rə̀-kə̀-ə̀ə̀*  
 plant-INCP-IRR-NMLZ:LOC/OBL-TOP  
 "It's this paddy here that's the one we're to plant tomorrow" (Post 2007: 722)
- (c) *acín*                      *dólà(a)-zù*  
 cooked\_rice eat-IMP.SDIR-HORT.INCL  
 "Let's eat (a meal)"
- (d) *acín-ə̀ə̀m*                      *dótó-kée*  
 cooked\_rice-DOM<sub>2</sub> eat-IMP.ODIR-HORT.POL  
 "Eat the rice" (Post 2007: 725)

The distribution of object markers in Galo is summarised in Table (6.12):

<i>nè</i>	<i>-m</i>	<i>ə̀ə̀m</i>
proper names	1/2 pro, reflexives	III pers. sing pro, dual and plural pro
given human DOs	simple demonstratives	non-human definite DOs

Table 6.6: Distribution of object markers in Galo

Diachronic evidence points to topicality as the main triggering factor for DOM. As we have seen above, the object marker for definite entities has two allomorphs *-m* and *(-)ə̀ə̀m*. According to Post (2007: 721), the latter allomorph, *(-)ə̀ə̀m*, derives from an earlier demonstrative *\*ə̀ə̀* fused with the original object marker *\*m*, which has been preserved only with pronouns. Interestingly, the form *\*ə̀ə̀* still survives in Modern Galo, where it has been grammaticalised into a topic marker (Post 2007: 703)<sup>30</sup>. Indeed, Post himself (2007: 721) hypothesises that *ə̀ə̀m* reflects a pattern where DOM was used to overtly code only topical direct objects by adding the topic marker *\*ə̀ə̀* to the original accusative form. That this is the most plausible explanation for the emergence of DOM in Galo is proved by evidence about the synchronic distribution of the topic marker and the accusative forms. As a matter of fact, the topic marker can never co-occur with the accusative (Post 2007: 707), while it is found on subjects/agents, genitives, inalienably possessed entities, instrumentals, sources (Post 2007: 709-714). In addition, the marker *ə̀ə̀m* is used to encode some temporal phrases and some temporal subordinate clauses

<sup>30</sup>Cf. Post (2007: 703) "The principal function of Topic marker *ə̀ə̀* is marking of definite reference to an already well-individuated, identifiable and "accessible" (i.e. "known" or "given") entity, generally as it is construed as contrastive with other such entities for establishment as the topic of a given clause".

(Post 2007: 810) functioning as a frame for the event presented in the main clause (Post 2007: 811). This polysemy, as I have repeatedly observed, is quite frequent cross-linguistically.

The seemingly anomaly of the pronominal system, in which first and second singular pronouns retain the old marker, as opposed to third singular and plural/dual pronouns, can be readily accounted for if we consider the highly conservative nature of first and second pronouns, which are generally the most stable forms in the pronominal paradigm. As for the usage of the two object markers found in Galo, Post (2007: 727-729) states that, however, some mismatches in the distribution of the two markers are found, although the conditions governing these alternations are not yet clear. Thus, examples in which an animate direct object is overtly coded by *əəm* are found, as well as examples in which an inanimate direct object is coded by *-nè*. Interestingly, it is possible, albeit rare, to find double-coding of a direct object, regardless of its animacy. Post (2007: 728) explains these occurrences as a means to "forestall the possibility that some (probable highly animate and/or thematically important referent) is indeed being expressed in a non-subject function. The presence of the object marker found with common nouns thus seems a means to make the direct object stand out more than if it were coded only with the marker reserved to animate direct objects. In this respect, the presence of the marker appears to be still connected to topicality.

The diachronic data I have presented corroborates the topicality hypothesis I have defended so far, at least as far as the second DOM pattern is concerned. However, in the Galo DOM system, definiteness seems to be taking priority over information-structure related parameters through a process of extension.

Dolakha Newari (Genetti 1997, 2007), Burmese (Jenny 2009), and Chepang (Noonan 1991) are other three clear examples of the relevance of topicality and topicality-related parameters for the presence of DOM. Let us start by describing Genetti's findings on Dolakha Newari. In this language, the case marker *-ta* functions both as a dative and accusative marker. Genetti (1997: 57) summarises the conditions for the appearance of DOM as follows:

- Human patients are marked when they are given or accessible;
- Animate patients are marked when they are topical;
- Inanimate patients are never marked.

First and second pronoun direct objects are always overtly coded, since they are always given in discourse, as in (6.121a, b):

(6.121) Dolakha Newari (Sino-Tibetan, Tibeto-Burman)

- (a) *jan-ta pulis-ke da-yeŋ*  
 1SG-DOM police-ALL PROH-take  
 "Don't take me to the police"
- (b) *thjj-ta mepsin helā yer-eu*  
 1PL.INCL-DOM other.ERG insult do-FUT.2PL  
 "Others insult us" (Genetti 2007: 115)

For third-person referents, DOM appears if they have been mentioned in the previous discourse, as shown by examples (6.122a) and (6.122b). In (6.122a), the referent "son" is introduced in the first line and then is overtly coded in the next occurrence as a direct object (Genetti 2007: 115):

(6.122) Dolakha Newari (Sino-Tibetan, Tibeto-Burman)

- (a) *thi-mā ake-uri optecā ju ām tākku*  
 one-CLF son-IND small be that time  
 "Then (they had) one small son at that time"
- (b) *ām mucā-ta bābu-ri-n mucā ju-el-lāgin*  
 that child-DOM father-IND-ERG child be-NMLZ2-because  
*muryā-ku ta-ene*  
 lap-LOC put-PTCP  
 "Because he was a child, the father put his child on his lap"

Genetti clearly demonstrates that animacy alone cannot predict the occurrence of DOM, as both human and animate direct objects are often left uncoded, as shown by examples (6.123a) and (6.123b), in which the same NP *misā* shows up as overtly coded in the latter case and uncoded in the former:

(6.123) Dolakha Newari (Sino-Tibetan, Tibeto-Burman)

- (a) *āme bā-n mebu misā harāi*  
 3SG.GEN father-ERG other woman bring.3SG.PRS  
 "Her father brings another woman" (Genetti 1997: 38)
- (b) *ām misā-ta-ri tar-ju*  
 that woman-DOM-IND keep-3SG.PST  
 "He kept that woman/wife" (Genetti 2007: 381)

Genetti motivates the optionality of DOM with human and animate direct objects in terms of topicality factors, namely activation and givenness (Chafe 1987). Her statistics on Dolakha Newari narrative texts reveal that in fact topicality controls the distribution of DOM. Once a referent "has been introduced, commented

upon and integrated into the discourse" (Genetti 1997: 49), it is more likely to be overtly coded than an inactivated and new one. This is why the same direct object may be either overtly coded or uncoded.

Similarly, in Burmese, topicality is the major factor that underlies the distribution of DOM, as has been pointed out by Thurgood (1978), and Jenny (2009a, 2009b). The object marker *-ko* occurs obligatorily on pronouns (6.124a), while with other NP classes, its usage relies upon the topicality of the direct object, as shown by its optionality in (6.124b):

(6.124) Burmese (Sino-Tibetan, Tibeto-Burman)

- (a) *ʔənɔ θú \*(ko) ʔaiʔ tɛ*  
 1M 3.AT DOM like NFUT  
 "I like him/her"
- (b) *khwè-khəlè (ha) ʔɛʔ (ko) laiʔ kaiʔ tɛ*  
 dog-DEM SM chicken DOM follow bite NFUT  
 "The dog is chasing and biting the chicken" (Jenny 2009)
- (c) *kà tə-sì (ko) ʔənɔ tó ʔəkoun.lòun hɣà tɛ*  
 car one-CLF DOM 1M PL all hire NFUT  
 "We all hired a car" (Jenny 2009)

As Jenny (2009a) has demonstrated, DOM in Burmese is not predicted if only animacy or formal definiteness are taken into account, as it is possible to find animate or definite direct object uncoded or optionally overtly coded, as examples (6.124b, c) above show. Indeed, Jenny argues that the distribution of DOM is better predicted by topicality, since topical direct objects are consistently overtly coded, as opposed to focal direct objects, which disfavour the use of *ko*. As a matter of fact, the direct object "frog", after being introduced, becomes a topic and is therefore overtly coded, as shown by (6.125a). In (6.125b), the direct object is focal and therefore DOM is strongly disfavoured (Jenny 2009):

(6.125) Burmese (Sino-Tibetan, Tibeto-Burman)

- (a) [...] *lu-khəlè ha phà ɕa yìn ɕa yìn tə-kaun*  
 [...] person-DEM SM frog seek while seek while one-CLF  
*hmá mə-yá tɛ ʔəshòun ye-ʔain nà hma phà*  
 just NEG-get NFUT.ATTR end water-pond near LOC frog  
*tə-kaun ko θwà twé tɛ*  
 one-CLF DOM go find NFUT  
 "The boy, while looking for frogs for some time, in the end he found  
 a frog near a water pond"

- (b) *ba sà t̪hin lɛ ?- khauʔ.shwɛ pɛ \*ko sà t̪hin tɛ*  
 what eat DES Q noodles EXCL DOM eat DES NFUT  
 "What would you like to eat? -Noodles"

Thus far, we have seen that DOM in Burmese is reserved to topical direct objects and disfavoured with focal ones. This is confirmed by the fact that *ko* tends to appear when the direct object is topicalised in clause-initial position, while it is less used when the direct object is adjacent to the verb (Thurgood 1978). Further evidence for the topicality requirements comes from other usages of *ko*. Aside from being used with temporal and spatial expressions and as object marker, *ko* is used as a topic marker, as shown by (6.126), which is the reply to the question "How much is that book?":

(6.126) Burmese (Sino-Tibetan, Tibeto-Burman)

- di sa.ʔouʔ ko mə-yàun phù*  
 this book DOM NEG-sell NEG  
 "This book is not for sale" (Jenny 2009)

In point of fact, Thurgood (1978) claims that *ko* has a topic-shift function, being used when a direct object is re-introduced in the discourse. This complies with the general functions we have seen at work in other languages with DOM.

In Chantyal (Noonan 1991), DOM seems to depend upon the animacy/topicality of the direct object, similarly to Burmese and Dolakha Newari. Human direct objects are usually overtly coded, while with animate ones there is variation:

(6.127) Chantyal (Sino-Tibetan, Tibeto-Burman)

- (a) *na-sə ram-ye cɦame-ra mara-i*  
 1SG-ERG Ram-GEN daughter-DOM see-PST  
 "I visited Ram's daughter"
- (b) *kyata-sə cu nɦaka t<sup>h</sup>a-i*  
 boy-ERG this chicken cut-PST  
 "The boy killed this chicken"
- (c) *kyata-sə cu nɦaka-ra t<sup>h</sup>a-i*  
 boy-ERG this chicken-DOM cut-PST  
 "The boy cut this chicken" (Noonan 1991: 53)

According to Noonan (1991: 54), direct objects are generally overtly coded when they are topicalised. It appears that in this latter case even inanimate direct objects can receive DOM, as shown by examples (6.128):

## (6.128) Chantyal (Sino-Tibetan, Tibeto-Burman)

- (a) *dfunŋ-ra dɸias-si k<sup>h</sup>ə-w*  
 tree-DOM touch-PFV come-IMP  
 "Come touch the tree"
- (b) *mɸaar-ra p<sup>h</sup>alam-sə t<sup>h</sup>a-m*  
 gold-DOM iron-ERG cut-PRS  
 "Iron cuts gold (i.e. Gold is cut by iron)"

Similarly, in Chepang (Caughley 1982, Thompson 1990), DOM seems to be closely associated with topicality. As shown by examples (6.129a, b), DOM occurs when the direct object is topicalised in initial position, while the direct object stays uncoded when the standard SOV word order is followed. According to Næss (2007: 71), who follows Caughley (1982: 65), there is an alleged difference in terms of affectedness when the direct object is overtly coded. Thus, the direct object in (6.129a) stays uncoded because it is not intentionally affected, while the direct object in (6.129b) is.

## (6.129) Chepang (Sino-Tibetan, Tibeto-Burman)

- (a) *Puʔ-nis-ʔi həw sat-ʔaka-c-u*  
 older\_brother-DU-ERG younger\_brother kill-PST-DU-A  
 "The two older brothers killed the younger brother"
- (b) *həw-kay puʔ-nis-ʔi sat-ʔa-thəy*  
 younger\_brother-DOM older\_brother-DU-A kill-PST-OBJ  
 "The two older brothers killed the younger brother" (Caughley 1982: 68)

Matisoff (1973) discusses the functions of *thàʔ* in Lahu. DOM in Lahu has apparently no effect on the semantics of the clause, and its contribution seems to be more pragmatic in nature, as it is used when "special emphasis is desired" (Matisoff 1973: 55-56). Matisoff further asserts that *thàʔ* is generally left out when the direct object is adjacent to the verb and the standard SOV word order is followed. Similarly to many languages we have seen in this chapter, *thàʔ* occurs with other constituents that are all linked by frame-setting, topical function. Indeed, it is found on temporal and spatial expressions, and it seems to be likely that the preposition used to overtly code direct objects developed out of a locative form.



## 6.13 Conclusion

In this chapter I have investigated 100 languages DOM and DOI with regard to their link with topicality and topic-marking constructions. The survey conducted here has shown that the appearance and development of DOM strongly correlates with the topicality of the direct object. As a matter of fact, DOM and DOI are overwhelmingly associated with the given and identifiable status of the direct objects, features that are typical of topics. Therefore, in order to trigger the use of DOM and/or DOI, the referent of the direct object must have been introduced into the discourse and commented upon. Crucially, however, DOM and DOI differ as to the functions they fulfil in discourse. As I have widely argued throughout this work, DOM is associated with the signalling of topic discontinuities, such as topic shifts and topic promotions. As a matter of fact, in many languages, the referents of overtly coded direct objects are often less continuous and accessible, since they have been absent from the discourse for a while. More importantly, with more than chance frequency, all the direct objects introduced by DOM become the topic of the ensuing discourse, thus complying with the topic shift/promotion function I have hypothesised. By contrast, DOI is associated with the establishment and the maintenance of topic continuity, as it is exploited to index highly continuous and accessible referents. Both DOM and DOI, however, seem to be highly disfavoured with newly-introduced direct objects, which are not part of the universe of discourse yet. This restriction is reflected by the impossibility for indefinite (non)-specific direct objects -the cut-off points varying from language to language- to occur with DOM. Indefinite (non)-specific referents are indeed strongly disfavoured as topics.

The close link of DOM and DOI with information structure is reflected by the recurrent association of these constructions with specific morpho-syntactic constructions employed to promote an entity to topic position or shift the topic from an entity to another entity, namely dislocations and topicalisations. There is plenty of reliable evidence that the presence of DOM and DOI correlates with topic-encoding constructions and variations in word order dictated by information structure. I have also shown that, in many cases, the dislocation of the direct object renders the presence of DOM and/or DOI mandatory. Although this structural correlation might be interpreted as having to do with the need to keep arguments distinct within the clause, I believe that the synchronic and diachronic evidence presented in this chapter goes against this explanation. From a synchronic point of view, we have seen how in many languages there are other means for identifying correctly the arguments of the clause which would render the presence of DOM unnecessary from a distinguishing perspective.

From a diachronic point of view, I have shown that there is ample evidence for the primacy of topicality in the development of DOM systems. Throughout

this chapter, we have seen that, in a lot of languages, object markers are part of polysemy patterns involving (i) frame-setting expressions, such as spatio-temporal expressions (as in many Indo-European, Afro-Asiatic, Tucanoan, and Altaic languages), and (ii) topic and conditional markers (as in some Tibeto-Burman, Nilo-Saharan, and Austronesian languages). In some cases, even if the connection between object markers and topic markers is no longer transparent synchronically, it can still be reconstructed based on diachronic evidence, as we have seen for Galo. In addition, incipient systems as well as diachronic data clearly show that DOM and DOI systems typically originate from the grammaticalisation of topic constructions, and they involve animate and definite direct objects because these particular types of direct objects, as opposed to others, are more likely to function as topics. These cross-linguistic findings are further corroborated by the analysis of diachronic comparative data from Romance languages I will present in Chapter 8.

Unlike Dalrymple and Nikolaeva's (2011) proposal, according to which the connection between topicality and DOM/DOI is found at the level of sentence-topicality, my analysis shows that the influence of topicality on DOM and DOI must be seen in the broader context of the discourse or the text, rather than as a notion limited to the sentence-level.

As a matter of fact, we have seen that the sentence-level and the discourse/text level somehow overlap when we look at the functions of DOM and DOI. At the sentence level, overtly coded direct objects are unquestionably topical, and are often primary topics. In addition, the presence of DOM must be seen as a strategy that serves to signal topic alternations and changes throughout the discourse, as is made clear by the regular association between the presence of DOM and topic-shifting/topic-promoting devices.

# Chapter 7

## DOM in serial verb constructions

### 7.1 Introduction

In this chapter, I will investigate the connection between DOM and serial verb constructions. In many languages, of which Mandarin Chinese is perhaps the most well known example, object markers diachronically emerge from serial verb constructions. As we will see in this survey, DOM in these languages is often only one of the functions covered by the construction. Furthermore, in other languages, the serial verb construction still retains its verbal properties, thus showing a dual behaviour.

The constructions that will be analysed in this chapter can be represented in general terms as in (7.1):

(7.1) (AGENT) - OBJECT MARKER + DIRECT OBJECT -V

The object marker usually derives from a (former) verb occurring as the first verb in a serial verb construction. The meanings of such verbs are quite diverse, as well as the functions they have acquired. As for the former, in many languages the object marker can be traced back to a verb meaning "hold, take, grasp". This grammaticalisation pattern has long been recognised in the literature (Lord 1982, 1993; Heine and Kuteva 2002) and is widespread in the West African languages studied by Lord (1993) as well as in Sinitic languages (Chinese being the most well-known instance). Chappell (2006, to appear) has identified other sources for object markers in Sinitic languages, most notably comitative constructions and verbs of giving and helping. These latter sources have been unrecognised so far in the literature and are particularly interesting insofar as they shed light on possible grammaticalisation paths that have gone virtually unnoticed (see discussion below).

All the languages surveyed in this chapter share the SOV schema given in (7.1) as far as DOM is concerned. This variation in word order is very remarkable, as all the languages surveyed in this chapter have a basic SVO word order. As I will discuss below, the differences in word order when DOM is present are a remnant of the original environment in which the construction arose. In addition, the preverbal vs. postverbal position of the direct object synchronically signals a difference in the information-structural status of the direct object. As a matter of fact, the preverbal position tends to be exploited for topical direct objects, while the postverbal one is typically occupied by focal, new direct objects.

In the following pages, I will therefore argue that, even in the languages where DOM derives from a serial verb construction, the distribution of DOM can be accounted for in terms of information structure parameters.

## 7.2 Sinitic

Sinitic languages are probably the most studied languages as far as the development of object markers out of serial verb constructions is concerned. Although the most well-known example is Mandarin Chinese, Sinitic languages show a diverse array of DOM constructions (called "disposal construction" in the specialised literature), which crucially differ from Mandarin Chinese both synchronically and diachronically, as brilliantly argued by Chappell (2006). I will start my survey by summarising the properties of DOM in Mandarin Chinese. Then, the distribution of the so-called "disposal constructions" in six non-Mandarin languages (Southern Min, Hakka, Cantonese, Shanghainese, Xiang and Gan) will be examined, based upon data gleaned from Chappell (2006, to appear).

As is well known, Mandarin Chinese has a construction, commonly referred to as "the *bǎ* construction" (Li and Thompson 1981: 463), in which the direct object is placed after *bǎ*, whose original meaning was "take" but before the verb. The use of this construction is regulated by a number of constraints. In order for a *bǎ*-sentence to be grammatical, the following three conditions must be fulfilled (Yip and Rimmington 2004: 200 ff.; Liu 2007, among others):

1. *bǎ* cannot be used with post-verbal direct objects;
2. *bǎ*-coded direct objects should have definite or specific referents;
3. the predicate must be complex, i.e. the main verb has to be followed by another constituent indicating boundedness (see below).

Animate and human direct objects show a strong tendency to be overtly coded by *bǎ*, although a human direct object can show up uncoded, as illustrated by

(7.2a) as opposed to (7.2b), in which the pronoun "him" is overtly coded in the former case and uncoded in the latter:

(7.2) Mandarin Chinese (Sino-Tibetan, Chinese)

- (a) *wo ba ta sha-le*  
 1SG DOM him kill-PFV  
 "I killed him"
- (b) *wo sha-le ta-(le)*  
 1SG kill-PFV him-PFV  
 "I killed him" (Li 2006: 377)

As for definiteness, the presence of *bǎ* makes available only a definite interpretation (7.3a), as opposed to uncoded direct objects, for which either a definite or an indefinite interpretation is available. As we have already pointed out above, uncoded direct objects must be in the default post-verbal position (7.3b):

(7.3) Mandarin Chinese (Sino-Tibetan, Chinese)

- (a) *wo ba juzi bo-le*  
 1SG DOM orange peel-PFV  
 "I peeled that orange"
- (b) *wo bo-le juzi*  
 1SG peel-PFV orange  
 "I peeled an orange" (Li 2006: 418)

A large and growing body of literature has investigated the *bǎ* construction, and many analyses have been proposed to account for its distribution (Li and Thompson 1981, Sun 1996, Li 1990, to name only a few). According to one influential line of research, the use of *bǎ* in Mandarin Chinese is conditioned by the degree of affectedness of the direct object. Indeed, sentences containing direct object *bǎ* have often been argued to express "disposal" (a calque of a Chinese term roughly meaning "affectedness"), i.e. "how a person is handled, manipulated, or dealt with: how something is disposed of; or how an affair is conducted" (Wang 1947, quoted in Jing-Schmidt 2005: 67). In this view, direct objects have to be highly affected by the event in order to be overtly coded (Li 2001; Li and Thompson 1981: 465). Disposal analyses have been later incorporated into the Transitivity Hypothesis proposed by Hopper and Thompson (1980). Within this framework, *bǎ* is analysed as a marker of high transitivity (Li and Thompson 1980, Sun 1996), since it occurs with definite direct objects and is subject to a number of aspectual constraints, such as event boundedness (Liu 1997). As a

matter of fact, *bǎ*-clauses in Mandarin Chinese require that the verb be followed by a constituent signalling even boundedness (i.e. telicity), such as the resultative marker *de*, the perfective marker *-le*, or a quantified phrase (Liu 1997)<sup>1</sup>. The fact that the event has to be bounded and the direct object definite has led many linguists to claim that what *bǎ* signals is the affectedness of the direct object, as advocated by Chao (1967), Li and Thompson (1981), and Sun (1996), among others. However, the idea that affectedness determines the appearance of *bǎ* does not hold either synchronically or diachronically. Rather, the distribution of DOM seems to be better explainable in information-structural terms.

In Iemmolo (accepted), and Arcodia and Iemmolo (submitted), I argue that the influence of affectedness, defined as the degree of specificity of a predicate as to the change undergone by a participant along a scale of possible changes (Beavers 2011, Croft to appear), seems to be null. For instance, no difference in affectedness can be detected in examples (7.2a, b) given above. The direct object "him" is governed by the verb "kill" in both examples. Although "kill" entails a complete, definite change of a definite, animate direct object in both cases, *bǎ* can be omitted, and the SVO counterpart is used instead, without bringing about any difference in the degree of affectedness of the direct object. Likewise, the direct object "orange" in (7.3a, b) above shows up either overtly coded or uncoded. In both cases, the orange has undergone a change of state, having been peeled off. Nonetheless, *bǎ* can be omitted and the SVO counterpart is then used.

Further evidence for the negligible role of affectedness comes from the fact that *bǎ* is often found with verbs entailing low affectedness or no affectedness at all, like stative or psychological predicates, as in (7.4):

(7.4) Mandarin Chinese (Sino-Tibetan, Chinese)

- (a) *wo ba ta de mingzi wangji-le*  
 1SG DOM 3SG DET name forget-PFV  
 "I forgot her name" (Arcodia and Iemmolo submitted)
- (b) *wo ba zhe-juzi nian-le san-xiaoshi-le*  
 1SG DOM this-sentence read-PFV three-hour-PRT  
 "I read this sentence for three hours" (Li 2006: 424)
- (c) *ta ba ni xiang de fan dou bu ken chi*  
 3SG DOM 2SG miss RES food all not will eat

<sup>1</sup>Liu (1997), however, does not consider this latter constraint as conclusive evidence for a close connection between affectedness and the presence of *bǎ*. Indeed, it should be further noted that the event boundedness hypothesis is challenged by the fact that *bǎ* is compatible with the durative atelic marker *zhe* (Jing-Schmidt 2005: 167). In addition, as Peyraube (1985: 195) observes, a verb alone may be found in final position if it is bisyllabic.

"S/he misses you so much that s/he won't even eat her/his meals" (Li and Thompson 1981: 469)

In order to account for the presence of *bǎ* in these examples, it has been suggested that the affectedness need not be physical, but rather psychological or imaginary (such as with emotional or subjective changes, see Li and Thompson 1981: 469-470). For instance, Li and Thompson (1981: 470) argue that the use of *bǎ* in (7.4c) is due to the addition of the post-verbal modifier, which conveys a sense or an implication of affectedness along with the verb. This very loose notion of affectedness, which includes emotional and somehow subjective changes, would explain the presence of *bǎ* with stative and psychological predicates. A similar analysis can be applied to examples (7.4a, b), with the difference that there is no post-verbal intensifier. Nonetheless, *bǎ* appears in both examples. Thus, a view that takes affectedness as the main trigger for DOM in Mandarin Chinese fails to account for the many instances where the direct object is not affected at all by the event expressed by the predicate.

What then triggers the use of the *bǎ*-construction? The distribution of DOM in Mandarin Chinese becomes less peculiar when cross-linguistic data from similar systems and information-structure parameters are taken into account. Indeed, DOM in Mandarin Chinese, as well as in other languages where object markers derive from former serial verb constructions, seems to be, once again, connected with topicality. For instance, Tsao (1987) proposed that the *bǎ*-construction represents a sort of "secondary" topic within the clause (a notion that he does not define in the same way as Nikolaeva (2001) and Dalrymple and Nikolaeva (2011)).

Let us briefly examine the properties that corroborate the reliance of Mandarin Chinese DOM upon topicality. First of all, the use of *bǎ* in an SOV word order contrasts with a post-verbal direct object, as we have seen above, and a structure with a preposed direct object without *bǎ*. As for the latter construction, it has frequently been observed in the literature that it is often used along with additive and scalar particles like "also, even, only" or negation and carries a strong contrastive meaning (Sun and Givón 1985: 346)<sup>2</sup>. Post-verbal direct objects are instead focal (LaPolla 1993).

Overtly coded direct objects possess many properties usually associated with topics. They are overwhelmingly associated with given information and highly identifiable referents, as shown by the corpus study in Arcodia and Iemmolo (submitted). In addition, *bǎ* cannot be used with non-specific expressions, since they cannot be construed as topical. Interestingly, as Sun and Givón (1985) pointed out, the entire (S)OV category involves some kind of contrast/emphasis which is comparable to (contrastive) topicalisation (Sun and Givón 1985: 338). Similarly,

<sup>2</sup>Whether preposed direct objects should be considered as contrastive topics or foci is still an open question. See Paul (2005), Liu (2007), among others, for discussion.

LaPolla (1995: 310) argues that topical NPs occur preverbally, while non-topical NPs occur post-verbally<sup>3</sup>.

The connection of *bǎ* with topicality is corroborated by an apparently marginal usage. As a matter of fact, Tsao (1987: 17) gives some examples in which *bǎ* is employed with non-argument NPs fulfilling different functions, such as locative (7.5a), and relational adverbials (7.5b):

(7.5) Mandarin Chinese (Sino-Tibetan, Chinese)

(a) *ta ba bilu sheng-le huo*  
 3SG DOM fireplace build-PFV fire  
 "She built a fire in the fireplace"

(b) *Ta ba nei jian shi xie-le yi feng baogao*  
 3SG DOM that CLF matter write-PFV one CLF report  
 "What s/he did with (about) that matter was to write a report about it"  
 (Tsao 1987: 17-19)

The use of the object marker to signal spatial expressions, as well as "relational" adverbials lends further support to the hypothesis that topicality is a fundamental feature in DOM. As I have shown in Chapter 6, this polysemy pattern is recurrent cross-linguistically. Aside from these marginal usages, *bǎ* may also be found on possessors, as in (7.6a), or an NP that is in a part-whole relation, as in (7.6b):

(7.6) Mandarin Chinese (Sino-Tibetan, Chinese)

(a) *tufei ba ta sha-le fuqin*  
 bandits DOM 3SG kill-PFV father  
 "Bandits killed his father" (Li 2001: 403)

(b) *ta ba shuiguo chi-le yi-ban*  
 3SG DOM fruit eat-PFV one-half  
 "He ate one half of the fruit" (Li 2001: 385)

As we will see shortly, in all the languages where the object marker derives from a serial verb the presence of DOM invariably brings about a change in word order that can readily explained as a change in information structure comparable to that of Mandarin Chinese.

Let us now turn to the other non-Mandarin Sinitic languages investigated by Chappell (2006, to appear). Chappell (to appear) provides the following summary

<sup>3</sup>LaPolla has challenged the idea that subject and direct object are useful and viable linguistic categories when one is dealing with languages like Mandarin Chinese



of the possible construction types of DOM (disposal constructions, DC in the summary) found across Sinitic languages:

1. Common DC: (AGENT) - [OM + DO] -V
2. DC with resumptive pronoun: (AGENT) - OM +DO<sub>i</sub> -V1 -(V2)- PRONOUN<sub>i</sub>
3. DC with initial object followed by the OM and resumptive pronoun:  
DO<sub>i</sub> - [OM+PRONOUN<sub>i</sub>]- V
4. DC with clause-initial object followed by the OM and zero anaphora:  
DO - [OM+Ø] + V (adapted from Chappell to appear)

The first type is found in Mandarin and is the most common in languages where DOM has developed out of serial verb constructions. The second type is found in Hakka and Cantonese Yue (see below), while the third kind is found in Taiwanese Southern Min and Wenzhou dialect. The last construction will not be surveyed here, and is found in Archaic Chinese and some dialects.

As I have already mentioned, all these constructions share the pre-verbal (topical) position. Furthermore, in the second and third structure, there is an anaphoric pronoun that can be interpreted diachronically as a remnant of the former serial verb construction from which these structure arose (Peyraube 1996). This pronoun must necessarily be co-referential with the direct object, i.e. it must be anaphoric. I will begin by summarising the distribution of DOM in Taiwanese Southern Min, which employs the marker *kāng* -*kā* to overtly code pre-verbal, referential direct objects, similarly to Mandarin *bǎ*. Some examples are given in (7.7). Nonetheless, there are some differences in usage between Mandarin Chinese and Taiwanese Southern Min, discussed by Chappell (2006: 452):

(7.7) Southern Min (Sino-Tibetan, Chinese)

- (a) *só-í gún lóng kā k'ò t'ng-k'í-lâi*  
therefore 1PL all DOM trousers take\_off-DIR  
"So we all took our trousers off (to go swimming)"
- (b) *a lì kā lí ê khui-lat lóng iòng-khì a*  
PRT 2SG DOM 2SG GEN strenght all use-DIR PRT  
"You used up all your strenght" (Chappell 2006: 453)

Interestingly, in Southern Min the direct object can be found in initial, topical position, as in (7.8), followed by *kāh*, which is the result of the fusion of the object marker *kā* with the following third singular anaphoric pronoun *yī*. This construction is usually avoided in Standard Mandarin:

## (7.8) Southern Min (Sino-Tibetan, Chinese)

*m̃ng kah            kui<sup>n</sup>    kui<sup>n</sup>    khì-lâi a*  
 door DOM:3SG close close INCH PRT  
 "(We) closed, closed the door" (Chappell 2006: 454)

Hakka dialects use the marker *tsiong*<sup>44</sup>, a cognate of Medieval and literary Mandarin *jiāng* "guide, lead", which was another Mandarin object marker later superseded by *bǎ*. As Chappell (2006: 456) observes, *tsiong*<sup>44</sup> can be used either as in Mandarin (aside from the different marker) or it can appear with a post-verbal anaphoric pronoun:

## (7.9) Sung Him Tong Hakka and North-Eastern Hakka (Sino-Tibetan, Chinese)

(a) *tsut<sup>5</sup>-tsji<sup>1</sup>,    tsj'iu<sup>4</sup>    tsjiong<sup>1</sup>    ngjia<sup>3</sup>    tsak<sup>5</sup>    tsjiau<sup>2</sup>-tsai<sup>3</sup>    kjiu<sup>4</sup>-hoi<sup>1</sup>*  
 in\_the\_end then    DOM    that    CLF    child-DIM    save-PFV  
*lɔ*  
 PRT  
 "(Sima Guang) saved the child"

(b) *tsiōng lì    tchâc kē            nā-loî    chǐt-p'êt    kî*  
 DOM    this    CLF    chicken    bring    eat-COMPL    3SG  
 "Eat up all this chicken" (De cette poule, n'en laissez rien) (Chappell 2006: 457)

Cantonese Yue makes use of *jeung*<sup>1</sup>, cognate with the Old Mandarin object marker *jiang* seen above, to overtly code topical, pre-verbal direct objects. In (7.10a), the direct object is in initial position, while in (7.10b) the direct object is resumed by a post-verbal third person pronoun, like in Hakka (Chappell 2006: 460):

## (7.10) Cantonese (Sino-Tibetan, Chinese)

(a) *jēung néih dábaahn-sihng    yāt-go    baakyepó*  
 DOM    2SG    dress\_up-become    one-CLF    old.lady  
 "Dress you up like an old lady"

(b) *Chìnkèih òh.hóu    jēung dī            tàuhfaat    yíhm-hāk    kéuih*  
 be.sure    NEG:IMP    DOM    CLF.PL    hair    dye-black    3SG  
 "Be sure not to dye your hair black" (Chappell 2006: 460)

Chappell (2006: 464) shows that in the Wu dialects the anaphoric pronoun is used along with DOM when the direct object is in initial position, as shown by (7.11a), from Wenzhou dialect, (7.11b) from Shaoxing<sup>4</sup>:

(7.11) Chinese Wu (Sino-Tibetan, Chinese)

(a) *beŋ<sup>31</sup>ku<sup>35</sup> dei<sup>11</sup> geu<sup>31</sup> ts<sup>h</sup>i<sup>313</sup> fiu<sup>o</sup>*  
 apple      DOM 3SG    eat      PRT  
 "Eat up the apple"

(b) *toŋ<sup>53</sup>ci<sup>53</sup> fi<sup>13</sup> tse<sup>ʔ<sup>45</sup></sup> ŋo<sup>13</sup>      ŋie<sup>n13</sup> p<sup>h</sup>a<sup>33</sup> dze<sup>o</sup>*  
 thing      3SG DOM 1SG.GEN do      broken PRT  
 "S/he broke my things" (Chappell 2006: 465)

In the Xiang dialect, an object marker cognate of Mandarin *bǎ* is employed, *pa<sup>41</sup>*, as in (7.12):

(7.12) Xiang (Sino-Tibetan, Chinese)

*pa<sup>41</sup> tɕ<sup>h</sup>yan<sup>41</sup>fu ta<sup>41</sup>k<sup>h</sup>at<sup>33</sup>*  
 DOM window    strike\_open  
 "Open the window!" (Chappell 2006: 466)

In Gan dialects, while there is not enough data to give a full account of the use of the object marker, DOM seems to be different than in Mandarin, as it allows monosyllabic verbs without any overt modification, unlike Mandarin Chinese (Chappell 2006: 469):

(7.13) Gan (Sino-Tibetan, Chinese)

*pa<sup>3</sup> cie<sup>3</sup> tsu<sup>3</sup>*  
 DOM 3SG boil  
 "Boil it!" (Chappell 2006: 469)

Finally, interesting data comes from a variety of Chinese spoken in the Qinghai province, Huangshui Chinese (Dede 2007). Huangshui Chinese employs the same marker as Mandarin *pa* to overtly code preverbal, topical direct objects, as in (7.14a):

<sup>4</sup>That a link between the use of DOM and topicality does exist seems indirectly confirmed by Chappell's (2006: 464) statement that studies upon Chinese dialects often claim that "topicalised pre-verbal objects are more frequent than the use of the disposal construction".

## (7.14) Huangshui Chinese (Sino-Tibetan, Chinese)

*pa*<sup>3</sup> *t'a*<sup>1</sup> *ɬiɔ*<sup>4</sup> *lɛ*<sup>2</sup>  
 DOM 3SG call come  
 "Call him here" (Dede 2007: 868)

Huangshui Chinese employs another marker *xa*, which can be utilised either as an alternative to *pa* or in addition to it, as shown by examples (7.15a, b). Unlike *pa*, *xa* occurs after the direct object:

## (7.15) Huangshui Chinese (Sino-Tibetan, Chinese)

- (a) *nɔ*<sup>1</sup> *kɛ*<sup>1</sup>*fɪ*<sup>3</sup> *xa* *xu*<sup>1</sup> *liɔ*  
 1SG boiled water DOM drink PRT  
 "I drank the boiled water"
- (b) *ɬia*<sup>1</sup> *pa*<sup>3</sup> *mə*<sup>2</sup> *xa* *kuā* *ʂɔ* *liɔ*  
 3SG DOM door DOM close up PRT  
 "He closed the door" (Dede 2007: 866-867)

According to Dede (2007), the object marker *xa* is employed with highly topical direct objects, both in sentence-initial and sentence-medial (i.e. preverbal) position, as in (7.16a, b)<sup>5</sup>:

## (7.16) Huangshui Chinese (Sino-Tibetan, Chinese)

- (a) *zu*<sup>4</sup> *xa* *ku*<sup>3</sup>*wa* *tʂ*<sup>1</sup> *ʂɔ* *liɔ*  
 meat DOM dog eat up PFV  
 "The meat was eaten by the dog"
- (b) *nɔ*<sup>3</sup> *ni*<sup>3</sup> *xa* *tɔ̃* *j*<sup>1</sup>*kua*  
 1SG 2SG DOM wait a\_bit  
 "I'll wait for you a bit" (Dede 2007: 868)

Interestingly, the marker *xa* has another function in some sub-dialects of Huangshui Chinese (Huangyuan and parts of Huangzhong), namely that of conditional and topic marker, while in other areas (such as Xining) the topic marker is *xɔ*, being thus phonetically similar but not identical to the object marker (Dede 2007: 875-877). The topic marking usage is exemplified by (7.17):

<sup>5</sup>Dede states that the passive translation of (7.16a) only attempts at showing the highly topical nature of the direct object. Furthermore, *xa* can be found with indirect objects of ditransitive verbs (Dede 2007: 869).

(7.17) Huangshui Chinese (Sino-Tibetan, Chinese)

*nɔ<sup>1</sup> tɕ'j<sup>4</sup> xa pɕ<sup>3</sup> ni<sup>3</sup> tɕ'j<sup>4</sup> ts' fɕ<sup>1</sup>piã*  
 1SG go TOP COMP 2SG go PRT convenient

"It is more convenient for me to go than you" (Dede 2007: 876)

Dede hypothesises that *xa*, from the original topic marking function, was later extended to overtly code highly topical direct objects. The data analysed by Dede are noteworthy, as they provide additional evidence for the relevance of topicality in determining the appearance of DOM. The extension of the topic marker to DOM is indeed a clear indicator that the condition that triggers the overt coding of direct objects is topicality, similarly to what we have observed in other languages thus far.

We have seen that in all Sinitic languages, the direct object is in pre-verbal position whenever the object marker is employed. As we will see at the end of this chapter, this requirement can be easily understood in information structure terms. Before drawing some conclusion on this pattern, I will present some data on Benue-Congo languages.

### 7.3 Benue-Congo

Many Benue-Congo languages of West Africa (Niger-Congo) display an object marking pattern strikingly similar to the one found in Sinitic. In all of these languages, when the direct object is topical, it receives the object marker and the word order changes from SVO to SOV, as we have seen in Sinitic languages. Let us start with Baule (Creissels 2010; Creissels and Kouadio 2010). Topical, definite direct objects in Baule may be overtly coded by *fa* "derived from a verb of taking", in what Creissels (2010: 7) remarkably calls "object fronting construction". A resumptive pronoun may optionally be present, as shown by (7.18b), as opposed to (7.18a):

(7.18) Baule (Niger-Congo, Kwa)

(a) *B'à kùn ákɔ'n*  
 3PL.PFV kill chicken.DEF  
 "They have killed the chicken"

(b) *B'à fã ákɔ'n b'à kùn í*  
 3PL-PFV take/DOM chicken.DEF 3PL.PFV kill 3SG  
 "They killed the chicken"

- (c) *B'à kà sikǎ'n*  
 3PL.PFV count money.DEF  
 "They have counted the money"
- (d) *B'à fà sikǎ'n b'à kà*  
 3PL.PFV take/DOM money.DEF 3PL.PFV count  
 "They have counted the money" (Creissels 2010: 7)

Creissels (2010: 7) argues that this construction is less grammaticalised than in other Benue-Congo languages, as the use of *fa* still entails physical contact. Therefore, (7.19b) is impossible because one does not take a snake to kill it, as opposed to (7.18b) above, where the use of *fa* is allowed since this situation is normal:

(7.19) Baule (Niger-Congo, Kwa)

- (a) *B'à k'ùn wǒ'n*  
 3PL.PFV kill snake.DEF  
 "They have killed the snake"
- (b) \**B'à f'à wǒ'n b'à k'ùn í*  
 3PL.PFV take snake.DEF 3PL.PFV kill 3SG  
 "They have killed the snake"

A further restriction reported by Creissels (2010) and Creissels and Kouadio (2010) concerns question and negative words, which cannot either be moved in pre-verbal position or be overtly coded by *fa*. Since question words and negation are generally associated with focality, this constraint gives further proof of the fact that this construction is reserved to topical direct objects.

Similar constraints connected with topicality are operative in other Benue-Congo languages where DOM is more grammaticalised. In Akan, *de*, which codes instruments and comitatives besides direct objects, is obligatorily employed with topical direct objects in pre-verbal position, while indefinite, non-topical direct object are uncoded and follow the verb (Lord 1993: 111-112), as shown by (7.20a) vs. (7.20b):

(7.20) Akan (Niger-Congo, Kwa)

- (a) *ɔ-de siká nó maa me*  
 3SG-DOM money DEF gave me  
 "He gave me the money"

- (b) \*ɔ-*maa me siká nó*  
 3SG-gave me money DEF  
 "He gave me the money" (Lord 1993: 112)

According to Lord (1993: 117) *kè* in Ga is being re-analysed as an object marker. Originally a verb of taking, *kè* developed into a marker of instruments and comitatives. As usual, it is found with topical, pre-verbal direct objects, as in (7.21):

(7.21) Ga (Niger-Congo, Kwa)

- (a) *è kè nù wò tó lè mlĩ*  
 3SG.F DOM water put bottle the inside  
 "She put water in the bottle"
- (b) *è wò tó lè mlĩ nù*  
 3SG.F put bottle the inside water  
 "She put water in the bottle" (Lord 1993: 119)

The pattern found in Idoma is identical. The prefix *l-* codes definite direct objects in pre-verbal position, as in (7.22):

(7.22) Idoma (Niger-Congo, Idomoid)

- (a) *ó l-ɔcí má*  
 3SG DOM-tree saw  
 "S/he saw the tree"
- (b) *ó má 'ɔcí*  
 3SG.F saw tree  
 "S/he saw the tree" (Lord 1993: 122)

Güldemann (2007: 92) discusses the case of Nupe, where there is a marker *a* which apparently signals perfectivity, and derives from a former verb of taking. This marker appears when the direct object is pre-verbal, as in (7.23a) vs. (7.23b):

(7.23) Nupe (Niger-Congo, Nupoid)

- (a) *Musa á tsùkũ zũ*  
 Musa PFV stick break  
 "Musa has broken the stick"
- (b) *Musa zũ tsùkũ*  
 Musa break stick  
 "Musa broke the stick" (Güldemann 2007: 92)

Crucially, Güldemann (2007: 92-93) argues that the two sentences differ as far as their information structure is concerned. Whilst (7.23b) would be the answer to a question like "What did Musa break?", in which the direct object is the focus, (7.23a), with the SOV word order and the *á*, would be the response to the question "What did Musa do with (or to) the stick?", where the direct object is presupposed and therefore topical. In the SOV word order, which invariably correlates with DOM, it is the predicate the asserted, hence focal, portion of the sentence.

## 7.4 Conclusion

As I have briefly discussed above, both in Sinitic and Benue-Congo languages overtly coded direct objects must be in pre-verbal position. In turn, the pre-verbal position is strongly associated with presupposed, topical direct objects in both families. I will now draw some conclusions upon the patterns which relate DOM, once again, to information structure.

Indeed, word order in Chinese has been held as driven by the notions of topicality and focality. LaPolla (1995) and LaPolla and Poa (2006) argue that the ordering of NPs within the clause is due to the topical vs. focal status of these elements. The fact that subjects/agents appear in pre-verbal position is due to their topical nature, while the tendency for direct objects to appear in post-verbal position is a reflex of their focal nature. The pre-verbal and specificity requirements for overtly coded direct objects, found in every languages with DOM or disposal construction, abide by this principle. If the presence of the object marker requires that the direct object be topical, then the direct object must be out of the focal domain. Moreover, it must necessarily be definite or at least specific, because indefinite non-specific referents are strongly dispreferred as topics. Likewise, Güldemann (2007: 101-102) contends that the ordering of the object with respect to the predicate in Benue-Congo languages serves to signal important differences with respect to information structure. In his view, the pre-verbal position of the direct object in Benue-Congo languages is associated with the topical, presupposed status of the direct object. By contrast, all other things being equal, the post-verbal position is associated with the focal, asserted status of the direct object along with the predicate. Since pre-verbal direct objects tend to be topical, it follows that they must be definite or at least specific.

Thus, it seems that, once again, the presence of object marking is connected to the information packaging of the clause. I am not claiming here, though, that the languages with object markers derived from serial verbs behave in the same manner as the languages I have investigated in Chapter 6.

DOM in serial verb languages has indeed quite a different origin, which should not be disregarded when these constructions are taken into account. The direct ob-



jects of verbs that end up grammaticalising into object marker are already placed before the second verb, in a serial verb construction such as (S)-V1-O1-V2-(O2), where the first verb (V1) is the future object marker. Crucially, the first direct object (O1) is shared by the two predicates. In addition, a second pronominal direct object can be found co-referential with the first direct object, as we have seen in Cantonese, Hakka, Southern Min, and Baule.

As has been demonstrated by Peyraube (1985, 1994) for Mandarin, and Lord (1993) for Benue-Congo languages, the first step in the reanalysis of V1 as an object marker involved semantic bleaching. The initial process was the juxtaposition of two sentences describing a sequence of two distinct events, one in which some entity takes some object and another one in which this object is further dealt with. Due to the temporal sequentiality between the two events (indeed, one needs to take something before using it), and the very general meaning of the verb (usually take, get), the V1 started being employed as a grammatical marker for the following NP, a sort of "introductory" device.

As a result, only the second predicate depicts an event, whereas the former first predicate is re-analysed as an object marker (Peyraube 1994 on Mandarin, Lord 1993 on Benue-Congo). Such a hypothesis would explain the word order shift from SVO to SOV when the object marker is used. This constraint can be seen as a remnant of the situation in which the DO was shared by the two juxtaposed clauses.

One can further speculate about how topicality came to be associated with these constructions. A likely scenario is one in which DOM came to be restricted to topical direct objects because the direct object in the construction resulting from grammaticalisation occupied a position generally associated with presupposed, topical elements. This scenario would fit the Mandarin data, as well as the Benue-Congo languages I have examined in section (7.3). Another possibility is that the construction itself started as a means to overtly code topics. This explanation would fit the languages in which the direct object is set off in initial position and the object marker follows it with a resumptive pronoun or zero anaphora, as in Hakka, Southern Min, Cantonese, and the Wu dialect of Wenzhou. It is very remarkable, however, the fact that DOM once more shows up in non-standard word orders to change the usual information packaging of the clause.

Before concluding this chapter, I would like to discuss the uncommon diachronic sources for object markers found in some Sinitic languages. Aside from the well-known sources of verbs of taking and holding, which are found both in Sinitic and Benue-Congo, we have already mentioned that Sinitic languages have developed object markers out of comitative markers and verbs of giving and helping. Recall also that this pattern is found in some creoles spoken in South-East Asia as well (see Chapter 5). As brilliantly demonstrated by Chappell (2000, 2006, to appear), Southern Min and Hakka dialects object markers derive from

comitative markers (in turn derived from verbs) through an intermediate stage as an oblique marker (with the vague semantics of "with respect to"). In addition, Chappell (2006, to appear) has shown that, in other languages, the object marker clearly derives from a verb of giving or helping via an intermediate beneficiary stage. As I have mentioned above, comitatives and verbs of giving and helping are extremely infrequent in the languages of the world as sources for object markers, verbs of taking and holding being by far the most common source cross-linguistically. In the Benue-Congo languages I am aware of, the source of object markers is indeed a verb of taking. Nonetheless, the "give" source is found in the geographically and genetically unrelated Tuu family of Southern Africa (Güldemann 2009, to appear), where the dative marker probably came about via the grammaticalisation of a former verb of giving.

# Chapter 8

## A case study: DOM and topicality in Romance

### 8.1 Introduction

In the Romance language family, many languages show DOM, including e.g. Spanish, Sardinian, Romanian, and a number of Southern Italian dialects. In this chapter, I will concentrate my attention on some less-studied Romance languages and varieties, such as Northern Italian, Gallo-Italian dialects, French varieties, Catalan, and Portuguese. Then I will discuss the rise and distribution of DOM in Sicilian and Spanish. While in Northern Italian, Gallo-Italian dialects, and the French varieties DOM is at an incipient stage, Catalan and Portuguese exhibit a more articulated system. In the remainder of this chapter, I will argue that topicality turns out to be the primary parameter for understanding and explaining DOM, because in these languages DOM is triggered by the topical status of the NP which fills the role of direct object. The role of topicality in the development of DOM in Romance languages is relevant also diachronically, as has been demonstrated by Pensado (1985) for Spanish, Iemmolo (2009) for Sicilian, among others.

### 8.2 DOM in Romance: some general remarks

The literature on the development and distribution of DOM in Romance is rather extensive and will not be dealt with here for reasons of space. Suffice it to say that, traditionally, DOM in Romance has been seen as a means to distinguish agents and direct objects when the latter display semantic properties usually associated with the former (Diez 1863, Meyer-Lübke 1900, Rohlf's 1969, 1971, Bossong 1985, 1991, to name only a few). As I have widely argued in the previous chapters, this view is not borne out by cross-linguistic data. As far as Romance languages

are concerned, this hypothesis clashes with the fact that, diachronically, DOM in Romance languages started from personal pronouns, the only nominal category which retained a case distinction in all Romance languages (see below). If the presence of overt coding on some direct objects is motivated by the need to assign the correct syntactic role to the NPs in the clause, it is very unlikely for such a marking to have begun with personal pronouns. Indeed, there is no need to overtly code a category that already carries an overt distinction (such as the opposition between *io* "I" and *me* "me" in Italian). These observations are far from being new, as they have been repeatedly made by several scholars (Berretta 2002, Nocentini 1985, Sornicola 2000).

The idea that the information status of the direct object may be the source of DOM systems has been repeatedly proposed by many scholars in Romance linguistics (Niculescu 1959, Pensado 1995, Laca 1987, Leonetti 2003), as I have mentioned in Chapter 2. In the following sections, I will show that incipient DOM in less-studied Romance languages emerges in topic-marking constructions, namely dislocations.

In the remainder of this chapter, I will set out to demonstrate that DOM is a strategy that emerges with personal pronouns in (mainly left) dislocation contexts. DOM systems in these languages are supposed to be motivated by the need to signal that the relevant direct objects are atypical at the information-structure level, insofar as they are primary topics. I will then show that the grammaticalisation of the relevant constructions can take two directions, as postulated by Darlymple and Nikolaeva (2011):

1. DOM may be extended to non-topical objects which share features of topic-worthiness, as in Modern Sicilian. In this way, the link with information structure is lost or at least weakened.
2. DOM may be restricted to topical objects only, as in Catalan.

### 8.3 Northern and Standard Italian and Gallo-Italian dialects

In the literature on the topic, DOM has been traditionally considered to be absent from both Standard and Northern Italian and Gallo-Italian dialects of Northern Italy (cf. Rohlfs 1971). In the last decade, however, some studies have convincingly demonstrated that DOM is also well attested in these Italian varieties and dialects (Berretta 2002, Nocentini 1985). The data from Northern Italian used for this study have been collected from various sources. Some examples are taken from Berretta (2002) and Iemmolo (2010), as well as from naturally occurring

spontaneous speech (many examples were overheard); others have been extracted from corpora of Contemporary Italian, such as the Corpus LIP<sup>1</sup>.

The data from Gallo-Italian dialects have been elicited through questionnaires (see Appendix D). In addition, some popular dialect comedies have been analysed in order to ascertain the presence of the phenomenon in the written language. The paucity of data have not allowed a statistical analysis of this construction, the use of which is restricted to spoken and informal registers of the language.

### 8.3.1 Northern and Standard Italian

The presence of DOM, which is traditionally called *prepositional accusative* in Romance linguistics, in Northern and Standard Italian was firstly noted, to the best of my knowledge, by Nocentini (1985), who observes that DOM was present even in his Tuscan variety, where DOM is usually held to be absent. Benincà (1988) touches upon DOM in Northern and Standard Italian, describing its usage as marginal. As a matter of fact, she claims that, in order for DOM to appear, the direct object has to be a dislocated (preferably first and second) singular pronoun, with a resumptive pronoun within the clause, thus deviating from the fairly loose (S)V0 basic word order of Italian. In an earlier paper, Benincà (1986) had suggested that DOM is present with third person pronouns and even proper names if the predicate is a psychological verb like *convincere* "to convince", *soddisfare* "to satisfy", etc. A few years later, Berretta (1989) began to investigate the distribution of DOM in Northern Italian, with particular regard to the socio-linguistic variation that is found in the usages of the construction.

As I will show shortly, Benincà and Berretta's characterisations are fundamentally accurate, insofar as the presence of DOM in Northern Italian correlates with the dislocated, pronominal status of the direct object. Indeed, Berretta (1989: 26) argues that topicality, expressed through dislocation, is the main parameter underlying the distribution of DOM in Northern and Standard Italian.

Let me now summarise the contexts in which DOM appears in Northern Italian<sup>2</sup>. DOM in Northern and Standard Italian is in fact quite limited. It is fundamentally restricted to first and second-person pronouns in dislocated position with an (optional) resumptive pronoun within the clause, as in (8.1, 8.2, 8.3, 8.4).

(8.1) Context A: "She said that she saw four people trying to pick the lock on her front door"

A	me,	non	(mi)	convince	questo
DOM	me	NEG	CLIT.1SG	convince.PRS.1SG	this

<sup>1</sup><http://badip.uni-graz.at/>

<sup>2</sup>This section is based on Iemmolo (2010b).

"This does not convince me" (Iemmolo 2010b: 249)

(8.2) *Ma a me colpisce una cosa...*  
 but DOM me strike.PRS.3SG one thing  
 "I'm struck by one thing" (Corpus LIP, Milan, ME88G)

(8.3) *A te non ti sopporto più*  
 DOM you NEG CLIT.2SG tolerate.PRS.1SG more  
 "I cannot stand you any longer" (Iemmolo 2010b: 249)

(8.4) *A voi vi ha spaventato molto la ultima*  
 DOM you CLIT.2PL have.3SG.PRS frighten.PTCP much the last  
*scossa?*  
 shake  
 "And did the last shake frighten you a lot?" (Radio interview, Radio  
 Popolare, 04-07-2009)

The topical status of the overtly coded direct object is demonstrated first by its being outside the scope of negation, as shown by (8.5a, b). In addition, the impossibility of omitting the preposition shows its grammaticalised status, as in (8.5a). That the main constraint is the left-dislocated position of the direct object pronoun is shown by the impossibility to overtly code post-verbal direct objects *in situ*, as in (8.5b):

(8.5) a *\*Me, non (mi) convince questo*  
 me NEG CLIT.1SG convince.PRS.1SG this  
 "This does not convince me" (Iemmolo 2010: 249)  
 b *\*Non (mi) convince a ME questo*  
 NEG CLIT.1SG convince.PRS.1SG DOM me this  
 "This does not convince me"

Actually, as is pointed out by Berretta, the preposition can be omitted, but only if the clause has a strongly contrastive meaning and the direct object is focal. For example, in *TE non sopporto più, non lei*, if the preposition were to be omitted, the resumptive clitic would be no longer possible, and the preposed NP would be a contrastive focus bearing prosodic stress (Benincà 1998, Rizzi 1997). However, this usage does not appear to be common in either Northern Italian or Standard Italian. All my informants, when asked to judge the acceptability of this construction, said that they felt the use of a cleft sentence to be more natural in this case, i.e. *È TE che non sopporto più, non lei!* "It is YOU that I can't stand anymore".

I have said above that DOM is fundamentally limited to first/second person pronouns. However, examples are found in which DOM introduces a third person pronoun (either singular or plural), or a proper name, as shown by (8.6). While these occurrences appear to be quite rare (indeed, the overwhelming majority of the examples I collected contain a first/second person pronoun), they show that the phenomenon is unstable, and suggest that DOM is extending downwards the animacy hierarchy, provided that the overtly coded objects are dislocated and thus topical:

- (8.6) (a) *A loro le aspettava Adone*  
 DOM them 3PL.F.OBJ wait.PST.3SG Adone  
 "Adone was waiting for them"
- (b) ?*A lei non la aspettano*  
 DOM her not 3SG.F.OBJ wait.PRS.3PL  
 "They don't wait for her" (Berretta 2002: 127)
- (c) *a me non mi vedono più, quelli di Telecom*  
 DOM me NEG CLIT.1SG see.PRS.3PL more those of Telecom  
 "Telecom will never see me again" (overheard)
- (d) *Volevo raccontar-vi una cosa... Forse a*  
 want.IPFV.1SG tell.INF-CLIT.2PL one thing... Maybe DOM  
*Emiliani divertirà*  
 Emiliani amuse.FUT.3SG  
 "I'd like to tell you something. Maybe Emiliani will find it amusing"  
 (Giovanna Botteri, TG3 13-01-2010)
- (e) *A me (mi) fa arrabbiare la sua arroganza*  
 DOM me CLIT.1SG makes get\_angry the his arrogance  
 "His arrogance makes me angry" (Iemmolo 2010b: 250)

Berretta (2002: 130) observed that DOM in left dislocations is also a device used to signal the beginning of a new conversational turn, and carries a topic shift function within discourse, in accordance with the role that left dislocation fulfils in Italian (Duranti and Ochs 1979) as well as cross-linguistically (see Givón 1983, Prince 1997 for English). The topic shift function, however, is carried out by the use of the stressed form of the personal pronoun (the only possible form in dislocated contexts), as in (8.7), which is B's reply to A's statement. In contrast,

the unstressed (clitic) counterpart would be felicitous if the topic is continuous, as in (8.8)<sup>3</sup>:

(8.7) Context: A: "It took a long time for them to call for that work! I'm very depressed..."

*A me, mi hanno chiamato subito*  
 DOM me CLIT.1SG have.PRS.3PL call.PTCP.PST immediately  
 "They called me immediately" (Iemmolo 2010b: 249)

(8.8) *Mi hanno chiamato subito*  
 CLIT.1SG have.PRS.3PL call.PTCP.PST immediately  
 "They called me immediately" (Iemmolo 2010b: 249)

Indeed, examples drawn from corpora of spoken Italian confirm that the topic shift function is quite prominent when DOM is present, as shown by (8.9), where the topic of the first sentence is "they", expressed only through indexation on the verb, since Italian is a pro-drop language. The topic shifts to "me" in the next sentence: such shift is indeed signalled by the presence of DOM on the dislocated pronoun:

(8.9) *sì però dico perché hanno mandato via*  
 yes but say.PRS.1SG why have.PRS.3PL send.PTCP.PST away  
*comandato gli altri in un altro ente e a me*  
 second.PTCP.PST the others in one other agency and DOM me  
*mi hanno fatto rimanere qui quando uno*  
 CLIT.1SG have.PRS.3PL make.PTCP.PST keep.INF here when one  
*chiede di andare via deve essere*  
 ask.PRS.3SG COMP go.INF away must.PRS.3SG be.INF  
*mandato in un altro ente*  
 send.PTCP.PST in one other agency  
 "Yes, but why did they relocate the others to another agency, keeping me

<sup>3</sup>Obviously, the topic shift function of the stressed form of the pronouns is not the only function that such pronouns perform in discourse. Nonetheless, this is one of the main functions of the opposition between stressed and unstressed pronouns in Italian. In terms of referential strength, unstressed forms are indeed used for referent continuity, while stressed forms are used for shifts, re-introductions, or focalisations (the latter without the object marker, of course). For instance, in (8.8) the direct object is continuous and represented by the clitic form of the pronoun. In contrast, in the example (8.7) there is a shift, marked by both the use of the stressed form and the object marker. Similarly, if the subject is continuous, it is expressed via inflection and there is pro-drop. When there is a shift, or a new subject is introduced, the stressed form of the pronoun or a full NP must be used.



here? When one asks to go away, he should be relocated to another local agency" (Corpus LIP, FC419D, Florence)

In example (8.10), taken from an Internet forum, the discussion is about men's trench coats, which are the main topic of most of the sentences (5 out of 7). At the end of the example, the primary topic switches to the speaker. This shift is signalled, again, by the use of DOM with the left dislocated pronoun:

- (8.10) *sinceramente lo trovo veramente un*  
 sincerely CLIT.3SG.M.OBJ find.PRS.1SG really a  
*abbigliamento da donna e stop; voi che dite?*  
 clothing for women and stop you what say.PRS.2PL  
*sbaglio io o forse è un po' esagerato??*  
 be.wrong.PRS.1SG I or maybe be.PRS.3SG a bit exaggerated  
*poi, se consideriamo tutto, il trench è nient'altro che*  
 then if consider.PRS.1PL all the trench is nothing.more than  
*un impermeabile, quindi se usato con la pioggia...*  
 a raincoat then if use.PTCT.PST.M with the rain...  
*insomma, l'uomo diventerebbe praticamente come la*  
 in.short the.man become.COND.PRS.3SG practically like the  
*donna. a me non riesce a convincere*  
 woman. DOM me NEG manage.PRS.3SG COMP convince.INF  
*assolutamente*  
 absolutely

"Actually, I find it (the men's trench coat) quite womanish, that's it. What do you think? Am I wrong or is it really a little bit over the top? Moreover, if I consider everything, the men's trench coat is nothing more than a raincoat. Then, if used when it is raining... In short, a man would become like a woman. As for me, this doesn't convince me at all" (Internet forum, Verona, <http://tinyurl.com/678v5cp>)

As we have mentioned above, the topic shift function associated with the use of DOM is even more evident when the opposition between unaccented pronouns and stressed pronouns plus DOM is taken into account. As is well known, unstressed pronouns are used when the referent is highly topical and continuous. (8.11a) would therefore be the normal answer to a question like *Perché non vai in macchina?* "Why don't you go by car?". By contrast, DOM, which must be used with the stressed series of pronouns, is used when there is a topic shift. (8.11b) would thus be the reply to an assertion like *Io vado in macchina* "I go by car" in a conversation:

(8.11) .

- (a) *Perché mi spaventa la nebbia*  
 because CLIT.1SG frighten.3SG.PRS the fog  
 "Because fog frightens me"
- (b) *Ah no, a me (mi) spaventa la nebbia*  
 ah no DOM me CLIT.1SG frighten.3SG.PRS the fog  
 "No, the fog frightens me" (adapted from Berretta 1989: 26)

As already hinted at by Berretta (1989), it is noteworthy that the DOM-coded dislocated direct object in all the examples discussed above can hardly be considered external to the sentence. Indeed, left dislocations in Italian do not necessarily exhibit a different intonation pattern (and often they do not), i.e. there is no pause between the dislocated constituent and the rest of the sentence. This pause is usually represented by a comma in written language; interestingly, in none of the examples quoted above is the comma necessary. Hence, there is no morpho-syntactic and prosodic evidence for the extra-clausal status for the dislocated element.

As observed in Iemmolo (2010b: 250), DOM is also possible in right-dislocation contexts, as in (8.12). These occurrences, however, are very scarce and sociolinguistically marked:

- (8.12) *come ci vedranno adesso, a noi?*  
 how CLIT.1PL see.3PL.FUT now DOM us  
 "How will they see us now? (Iemmolo 2010b: 250)

As a matter of fact, many speakers reject them as impossible or hardly acceptable, and I have not been able to find any examples either in corpora or in Internet forums<sup>4</sup>. When in post-verbal position, indeed, the pronominal direct object does not receive DOM because it is focal, as shown by (8.13c).

(8.13) Context: Meeting at a Voluntary Association, Milan.

- A: *Qualcuno fra i presenti è*  
 someone among the present.PL aux.PRS.3SG  
*convinto di questa cosa qui?*  
 convince.PTCP.PST of this thing here  
 "Is anyone here convinced of this thing?"

<sup>4</sup>Berretta (1989: 22) contends that DOM in post-verbal position is possible, albeit much more restricted than in left-dislocated position. That DOM can be found in right-dislocation contexts is beyond doubt. Interestingly, speakers reject examples like (8.12) as typical of Southern Italian varieties.

- B: *No, ritengo che non sia credibile*  
 no, think.PRS.1SG COMP NEG be.SUBJ.PRS.3SG believable  
 "No, it doesn't seem plausible"
- C: *In effetti non convince neanche ME*  
 in effect NEG convince.PRS.3SG even me  
 "In fact, this does not convince even me"

Recall from Chapter 4 that the focal domain may contain topical elements (Lambrecht 1994, Nikolaeva 2001). As a matter of fact, if the direct object is topical within the focus domain, the post-verbal pronoun could be overtly coded but the clitic resumptive pronoun must be necessarily present, as shown by (8.14):

- (8.14) *In effetti, non mi convince neanche a me*  
 in effect NEG CLIT.1SG convince.PRS.3SG even DOM me  
 "In fact, this does not convince even me"

It must be noted that the use of the construction with the overtly coded direct object plus the clitic seems to trigger a sort of contrastive nuance. In addition, as is clear from example (8.14), the overtly coded direct object can be modified by additive, scalar, and exclusive particles such as *(ne)-anche* "even, too", *solo* "only", *perfino* "even", provided that the direct objects are pronominal and topical. The presence of DOM in these cases could be explained by the fact that these particles can apply their meaning to the (generally contrastive) topic of the clause (König 2001: 755).

Further evidence for the topic shift function of DOM comes from the fact that, when the subject is overtly expressed, it is usually found in post-verbal position. As is well known, post-verbal subjects and agents in Italian are usually focused (Andorno 2003: 89, Berretta 2002: 132, Lambrecht 1994: 137). Hence, when both DOM and an overtly expressed subject are present, the direct object is the topical constituent, while the subject is focal and usually accented, as exemplified by (8.15a, b):

- (8.15) (a) *sai che a me m'ha deluso la Bretagna?*  
 know.2SG.PRS COMP DOM me CLIT.1SG-have.PRS.3SG  
 disappoint.PTCT.PST the Bretagne  
 "You know, Bretagne disappointed me!" (Berretta 2002: 144)
- (b) *A me preoccupa Torino: è una città difficile [...] A Cerami invece preoccupa Viterbo*  
 DOM me worry.3SG.PRS Turin be.3SG.PRS a city  
 difficult [...] DOM Cerami instead worry.PRS.3SG Viterbo

"As for me, I'm worried about Turin, as it's a difficult city. As for Cerami, he's worried about Viterbo instead" (Berretta 2002: 147; from an interview to Roberto Benigni)

Indeed, as already pointed out by Berretta, the simultaneous presence of DOM and an overtly expressed subject results in the association of both grammatical relations with non-default values as far as information structure is concerned, i.e. subject: focus, direct object: topic. Similarly to the cases in which an overtly expressed subject is absent -and thus arguably topical- the use of DOM brings about a topic shift even when both arguments are expressed. Thus, DOM in Northern and Standard Italian seems to have a topic-shift function, regardless of the presence or absence of a subject.

Lastly, I briefly touch upon the typology of verbs that most frequently take DOM in Northern and Standard Italian. Following previous observations by Berretta (2002), two main classes of verbs can be identified. The first class includes transitive psychological predicates, such as *convincere* "convince", *persuadere* "persuade", *preoccupare* "worry", *disturbare* "disturb". The second class includes causative verbs (*fare/lasciare* plus infinitive). Further verbs that do not fall into either of these categories are, e.g., *aspettare* "wait for", *vedere* "see", *chiamare* "call", *portare* "take".

The verb classes with which DOM seems to be more stable are the first two, i.e. transitive psychological and causative predicates. This is demonstrated by various facts. For verbs belonging to these two classes, DOM is obligatory only if the direct object is a dislocated first or second person pronoun. This, however, applies to the other verbs that do not belong to any of these categories. What differentiates the first two classes from the third is the fact that the former, but not the latter, allow the omission of the resumptive clitic, as Berretta (1989) and Benincà (1988) observed as well. The omission of the clitic, exemplified by (8.1, 8.2, 8.5, 8.10, and 8.15b) is probably suggestive of greater grammaticalisation of these constructions with these verbs, which appear to be the ones with which DOM started. The more grammaticalised nature of DOM with these verbs, reflected by the omission of the clitic, is in turn a good indicator of greater syntactic integration between the overtly coded direct object and the governing predicate. It must be noted that, while the omission of the clitic does not change either the information structure of the clause or its syntax, it does have a different socio-linguistic distribution. As a matter of fact, the forms without clitic are held as more formal than the forms in which the clitic is instead used. The rest of the verbs, with which DOM is less common, still requires the clitic be present.

Summing up, the incipient DOM system found in Northern and Standard Italian clearly shows that DOM arises in marked discourse configurations, where the

direct object is a highly topical element that generally needs to be re-established in discourse, thus performing a strong topic-shift function.

### 8.3.2 Gallo-Italian dialects

In Iemmolo (2010b) I argued that, contrary to what Berretta (1989, 2002) and Zamboni (1989) contend, DOM is present in Gallo-Italian dialects as well. The presence of DOM in some peripheral Gallo-Italian dialects has been noted by Rohlfs (1971), who provided some examples of DOM in the dialect spoken in Trieste (Friùli Venezia-Giulia, as in (8.16):

- (8.16) *A mi tratar-me in sta manera?*  
 DOM me treat.INF-CLIT.1SG in this way?  
 "But why treat me like this?" (Trieste; Rohlfs 1971: 63)

Aside from these examples, the common view in the literature is that DOM is usually absent in Gallo-Italian dialects<sup>5</sup>. As a matter of fact, it appears that the most likely scenario is one in which DOM emerged in these dialects as a result of intense contact with Italian. For practical reasons, I restricted the scope of my enquiry to the dialects spoken in Lombardy, Emilia-Romagna (only Modena), and Piedmont (Tortona and Viguzzolo). In order to ascertain the presence of DOM in these dialects, I used two questionnaires, written in Italian, containing twenty-five sentences with overtly coded direct objects, that the informants were asked to translate into their dialect (see Appendix D). The first questionnaire included only sentences in which a pronominal/nominal direct object was left-dislocated and governed by the same verbs found in the regional varieties of Italian. In the second questionnaire, all direct objects were either in the normal post-verbal position or right-dislocated. The questionnaire was submitted to fifteen people from different places in Lombardy. Interestingly, in none of the dialects investigated was the direct object overtly coded when it was in post-verbal or right-dislocated position. On the contrary, DOM is well attested with left-dislocated direct objects, provided that they are pronominal, as shown by the following examples from Ono S. Pietro (Lombardy), and Viguzzolo (Piedmont):

- (8.17) Ono S. Pietro (Brescia)

<sup>5</sup>Nonetheless, instances of DOM are sparsely found over the time in Gallo-Italian dialects as well, as Formentin (2004) and Parry (2003) have demonstrated based on historical evidence from Veneto and Liguria respectively. As a matter of fact, examples of DOM are found even in Dante's "Divina Commedia", as witnessed by the following example: *Lascia parlare a me, ch'i' ho concetto/ ciò che tu vuoi* "Let me speak, because I have conceived/ that which thou wishest", in *Inferno* XXVI, 73-74.

(a) *a me i ma ciama hemper*  
 DOM me 3SG.M.SUBJ CLIT.1SG call.PRS.3SG always  
 "He always calls me"

(b) *a otre i va ciama hemper*  
 DOM you.2PL 3SG.M.SUBJ CLIT.2PL call.PRS.3SG always  
 "They always call you"

## (8.18) Viguzzolo (Piedmont)

(a) *a mi a m ciamen dop*  
 DOM me 3PL.SUBJ CLIT.1SG call.PRS.3PL later  
 "They call me later"

(b) *a ti a t ciamen dop*  
 DOM you.2SG 3PL.SUBJ CLIT.2SG call.3PL.PRS later  
 "They call you later"

The presence of DOM is quite unstable as far as the classes of nominals are concerned. As a matter of fact, the only nominals that always take DOM in all the dialects surveyed here are first/and second personal pronouns. For instance, while in the dialect of Ono S. Pietro DOM seems to occur with other pronouns as well, in the dialect of Viguzzolo DOM is restricted to first and second singular personal pronouns only, as demonstrated by the ungrammaticality of (8.19):

(8.19) \**a lü a l ciamen dop*  
 DOM him 3PL.SUBJ 3SG.OBJ call.3PL.PRS later  
 "They call him later"

We have seen that the cut-off point of DOM with pronouns varies according to the variety considered, with the dialects spoken in Vailate, Ono S. Pietro, Villastrada, and Barbianello (Lombardy) allowing DOM with first/second singular and plural dislocated pronouns (even if, for plural pronouns, speakers usually tend to judge better the option without DOM). In some other varieties, as in the dialect of Viguzzolo, Tortona (Piedmont), Pavia (Lombardy), Modena and Castel San Giovanni (Emilia-Romagna), speakers tend to reject DOM with pronouns other than first and second singular. Examples (8.20), from Modena, (8.21) from Castel San Giovanni, and (8.22) from Tortona, show the impossibility of DOM to occur with third person pronouns as well as with plural pronouns:

## (8.20) Modena (Emilia-Romagna)

(a) *A me a m scocia, tot cal lavor*  
 DOM me 3SG.SUBJ CLIT.1SG annoy.PRS.3SG all that work  
*che*  
 here  
 "That work annoys me"

(b) \**A lo a l scocia, tot cal lavor*  
 DOM him 3PL.SUBJ 3SG.M.OBJ annoy.PRS.3SG all that work  
*che*  
 here  
 "That work annoys him"

## (8.21) Castel San Giovanni (Emilia-Romagna)

(a) *A mi, i m cunusan tüt*  
 DOM me 3PL.SUBJ CLIT.1SG know.PRS.3PL everybody  
 "Everybody knows me"

(b) \**A lalù i l cunusan tüt*  
 DOM him 3PL.SUBJ 3SG.M.OBJ know.PRS.3PL everybody  
 "Everybody knows him"

## (8.22) Tortona (Piedmont)

(a) *A mi, i m cunusan tuti*  
 DOM me 3PL.SUBJ CLIT.1SG know.PRS.3PL everybody  
 "Everybody knows me"

(b) \**A lu i l cunusan tuti*  
 DOM him 3PL.SUBJ 3SG.M.OBJ know.PRS.3PL everybody  
 "Everybody knows him"

In none of the varieties examined here is DOM possible with other kind of nominals, no matter whether they are dislocated or not, as shown by the examples in (8.23). (*al* in front of masculine kinship terms and other kind of nominals is the definite article in these dialects).

(8.23) (a) (\**A Luigi, i l'an vest*  
 DOM Luigi 3PL.SUBJ 3SG.M.OBJ-have.PRS.3PL see.PTCP.PST  
*in piazza stamateina*  
 in square this:morning  
 "They saw Luigi in the square this morning" (Modena)

- (b) (\*A) *papà l'et vest?*  
 DOM papa 3SG.M.OBJ-have.PRS.2SG see.PTCP.PST  
 "Have you seen dad?" (Castel San Giovanni)
- (c) *Al lader l'a est niü?*  
 the thief 3SG.M.OBJ-have.PRS.3SG see.PTCP.PST nobody  
 "Have nobody seen the thief?" (Vailate)
- (d) *Al piculei u l'a*  
 the little\_child 2SG.SUBJ 3SG.M.OBJ-have.PRS.3SG  
*spavintà fanda incsi!*  
 scare.PTCP.PST doing so  
 "You scared the little child by doing so!" (Tortona)

Since data elicited through questionnaires cannot be deemed fully reliable, as speakers might tend to translate based on the Italian original version, I looked at occurrences of DOM in written texts. I therefore selected some comedies written in local dialects. I used texts from Bergamo, Milan, Brescia, and Mantua (Lombardy). Unsurprisingly, quite a few instances of DOM are found in these texts too. In all instances, DOM codes a first/second singular personal pronoun in dislocated position. Interestingly, all the cases of DOM are within topic-shift contexts, as shown by example (8.24), in which DOM is used to shift the topic from the doctor to the speaker herself:

- (8.24) (A:) *Adès però mé g'ho de indà. I óter*  
 now but 1SG there\_have.PRS.1SG to go. the.PL other  
*i visite la settimana che è. Ciao*  
 3PL.OBJ visit.PRS.1SG the week that come.PRS.3SG. Bye  
*a töc*  
 to all  
 "Now I have to go. I will visit the others again next week. Bye to everybody"
- (B:) *Certo che a mé ol nòst dutùr al*  
 certain that DOM 1SG the our doctor 3SG.M.SUBJ  
*ma convince mia tàt*  
 CLIT.1SG.OBJ convince.PRS.3SG NEG so\_much  
 "Actually, our doctor doesn't convince me so much" (Bergamasco, from *Quater ciacole so l'era*, p. 15)

The change of topic, however, does not seem to be the only function of DOM in (8.24). These examples, analysed within the context of a dialogue, show another interesting property of DOM in Northern Italian and Gallo-Italian dialects,



which deserves a more thorough investigation. Indeed, it appears that we are dealing with a sort of turn-taking device, which the speaker uses to take up the floor. This turn-taking function is even more evident in example (8.25):

(8.25) (A:) *Aiuto, aiuto! A i è dré*  
 help help ASS CLIT.3PL.SUBJ come.PRS.3PL behind  
*i carabinieri*  
 the.PL police  
 "Help, help! The police is following us!"

(B:) *perché a gh'ét de i-ga pura? me*  
 why ASS there-have.PRS.2SG to have.INF-there fear? 1SG  
*ho copàt nigü, ho mia*  
 have.PRS.1SG kill.PTCP.PST nobody have.PRS.1SG NEG  
*robàt e ho sèmper pagàt i*  
 steal.PTCP.PST and have.PRS.1SG always pay.PTCP.PRS the.PL  
*tase*  
 taxes

"Why are you so afraid? I haven't killed anybody, I haven't stolen anything and I have always paid the taxes"

(C:) *a mé i pöl mia arestà-m per*  
 DOM 1SG 3PL.SUBJ can.PRS.3PL NEG arrest.INF-1SG.OBJ for  
*vagabondaggio perché me a g'ho öna*  
 vagrancy, because 1SG ASS there-have.PRS.1SG a  
*panchina fissa.*  
 bench fixed

"They can't arrest me for vagrancy, because I have a permanent bench" (Bergamasco, from *Öna storia de barbù*, p. 22)

In (8.25c), the speaker is taking up the floor by referring, through the use of the overtly coded dislocated direct object, to himself, thus shifting the topic. This situation complies with Duranti and Ochs' (1979) observations on the usage of left dislocations in Italian conversation. Indeed, as we have already argued, direct object referents introduced by left dislocations are a means to re-introduce a referent which has been mentioned in the prior discourse but is not the current topic of conversation. I believe that this turn-taking function is carried out by the combined use of DOM in left dislocation contexts. The use of DOM in these contexts has a twofold motivation. On the sentence-level, it serves to signal the highly topical nature of the direct object, which in all the cases we have seen so far is more topical than the agent. Thus, DOM encodes the fact that the information structure of the sentence does not follow the predicted pattern. The high topicality

of overtly coded direct objects has its correlates at the discourse level, in that DOM shifts the topic of conversation from the immediately preceding topic (usually the subject) to another one, which is a grammatical direct object.

I will not further elaborate on this issue here, for which a more detailed study is needed. Nevertheless, the Italian and Gallo-Italian data is interesting in two respects. First, its emergent status allows us to uncover the processes that bring about DOM. As a matter of fact, the data we have analysed adheres to the cross-linguistic distribution of DOM, according to which DOM arises in precise discourse configurations and is governed by information structural parameters. Second, this emerging pattern provides us with precious insights on the chief role of discourse in shaping morpho-syntactic structure. Indeed, it makes it clear that grammatical structures like DOM spring up from usage and carry out particular communicative functions that are later conventionalised into grammar (Hopper 1987). In the next section, I will show that the conditions that determine the appearance of DOM in Northern and Standard Italian and Gallo-Italian dialects hold for other Romance languages with incipient DOM systems as well.

## 8.4 DOM in Catalan, Portuguese, and French varieties

In this section, I will look at the presence and distribution of DOM in French varieties, Catalan, and Portuguese. DOM in these languages displays a restricted distribution, quite similar to the one found in Italian. As a matter of fact, the presence of DOM has gone virtually unnoticed in French, while the presence of DOM in Catalan has been often held as a "Castilianism" in prescriptive Catalan grammars (see Badia 1994). As will become clear in the following discussion, there is a striking similarity between these languages with respect to the constraints that govern the appearance of DOM. Indeed, topicality is the key parameter for DOM to come about in both French and Catalan, where DOM is almost invariably found in dislocated structures.

### 8.4.1 Catalan and Portuguese

Let us start by analysing the Catalan data. Standard Catalan is assumed to present DOM only in limited contexts, such as personal pronouns (8.26a), the universal quantifiers *tothom* and *tots*, "all" (8.26b), the relative pronoun *el qual* (8.26c), and when the object pronoun appears in a reciprocal construction (8.26d). However, DOM is obligatory only with personal pronouns and the universal quantifiers, being optional in the other contexts (Escandell-Vidal 2009: 837-839):

- (8.26) (a) *Jo t'ajudo a ti i tu*  
 I CLIT.2SG\_help.PRS.1SG DOM you and you  
*m'ajudaras a mi'*  
 CLIT.1SG-help.FUT.2SG DOM me  
 "I help you and you will help me"
- (b) *Hi he saludat (a) tothom*  
 there have.1SG.PRS greet.PTCP.PST DOM everyone  
 "I greeted everyone"
- (c) *Vaig veure-hi el teu amic, el/al qual*  
 go.1SG.PRS see.INF-there the your friend, the/DOM.the which  
*volient presenter com a candidat*  
 want.3PL.PST present as a candidate  
 "There I saw your friend, whom they wanted to present as a candidate"
- (d) *Ens miràvem l'un a l'altre*  
 1PL.OBJ look.1PL.PST the\_one DOM the\_other  
 "We looked at each other" (Escandell-Vidal 2009: 838)

The distribution of DOM varies depending on diatopic and diastratic factors. DOM in the spoken language is also found with other lexical human direct objects, provided that they are dislocated, as shown by (8.27a, b) and (8.28a, b):

- (8.27) (a) *A Núria, no crec que la*  
 DOM Nuria NEG think.1SG.PRS that 3SG.F.OBJ  
*pugues convencer*  
 can.SUBJ.PRS.2SG persuade.INF  
 "Nuria, I don't think you can persuade her"
- (b) *Als funcionaris no els satisfà la*  
 DOM.the civil\_servants NEG 3PL.OBJ satisfy.3SG.PRS the  
*proposta*  
 proposal  
 "Civil servants are not satisfied by the proposal" (Escandell-Vidal 2009: 840)
- (8.28) (a) *A ta mare, la vaig vore ahir*  
 DOM your mother 3SG.F.OBJ go.PRS.1SG see.INF yesterday
- (b) *Vaig vore \*(a) ta mare*  
 go.PRS.1SG see.INF DOM your mother  
 "Your mother, I saw her yesterday" (Escandell-Vidal 2007: 31)

The presence of DOM in Catalan with dislocated direct objects was already observed by Rohlfs (1971: 70 ff.), who discussed at length many uses of DOM "pour mettre l'objet en relief" (lit. "to put the object into relief", i.e. to emphasise it) and cited the examples in (8.29)<sup>6</sup>:

- (8.29) (a) *a aquesta senyora, no la conec*  
 DOM this lady NEG 3SG.F.OBJ know.PRS.1SG  
 "This lady, I don't know her"
- (b) *an es pobres, Diu els ajuda*  
 DOM the poors God 3PL.OBJ help.PRS.3SG  
 "The poor, God helps them" (Rohlfs 1971: 70)

In a detailed analysis of DOM in Catalan, Escandell-Vidal (2007, 2009) shows that the distribution of DOM in the Catalan variety spoken in the Balearic Islands depends upon the topical status of the direct object. Aside from the contexts in which DOM is obligatory in Standard Catalan as well, like pronouns, universal quantifiers, etc., DOM in Balearic Catalan has been extended to every dislocated direct object, as long as they are definite or at least specific. Hence, even inanimate direct objects may be overtly coded, as in (8.30):

- (8.30) (a) *A aquesta darrera (frase) noltros la diríem així*  
 DOM this last sentence 1PL 3SG.F.OBJ  
 say.COND.PRS.1PL like\_that  
 "This last sentence, we would say (it, GI), like this"
- (b) *an es ganivets elsi vaig ficar an es calaix*  
 DOM the knives 3PL.OBJ go.PRS.1SG put.INF to the  
 drawer  
 "(As for) the knives, I put them in the drawer" (Escandell-Vidal 2009: 855)
- (c) *A ses pomes, mengemmos les*  
 DOM the apples eat.PRS.1PL 3PL.F.OBJ  
 "As for the apples, let's eat them" (Rohlfs 1971: 70)

The definiteness constraint appears to be motivated by the topical status of the dislocated constituents. However, Escandell-Vidal (2007, 2009) notes that

<sup>6</sup>The preposition *a* has two allomorphs, [ən] before an unstressed vowel and [ə] before a stressed vowel or a consonant (Escandell-Vidal 2009).

examples with indefinite dislocated direct objects, such as the one in (8.31), are accepted by some speakers. This does not come as a surprise, since this kind of NP must be interpreted as having a generic referent:

- (8.31) *A un gelat, me lo prendria amb molt de gust*  
 DOM a ice-cream, 1SG 3SG.M.OBJ take.PRS.COND.3SG with  
 much of pleasure  
 "As for an ice-cream, I'd really love to have one" (Escandell-Vidal 2007: 33)

DOM is also well attested in right dislocations, as shown by (8.32):

- (8.32) (a) *Jo la vaig llegir an aquesta (comedia)*  
 I 3SG.F.OBJ go.PRS.1SG read.INF DOM this comedy  
 "This comedy I've read it"  
 (b) *Les he trobades a ses faltes*  
 3PL.F.OBJ have.PRS.1SG find.PTCP.PST DOM the mistakes  
 "The mistakes, I've FOUND (them)" (Escandell-Vidal 2009: 847)

In all the examples discussed above, DOM would be ungrammatical if the direct object were in its canonical position, i.e. post-verbal, which is the dedicated focus position in Catalan (Vallduví 1992), as shown by the examples in (8.33), where the direct object *in situ* is not overtly coded, as opposed to the dislocated ones:

- (8.33) (a) *No estima en Joan*  
 NEG loves the Joan  
 "S/he does not love Joan"  
 (b) *An en Joan no l'estima*  
 DOM the Joan NEG 3SG.M.OBJ-love.3SG.PRS  
 "As for Joan, she/he does not love him"  
 (c) *No l'estima, an en Joan*  
 NEG 3SG.M.OBJ-love.3SG.PRS DOM the Joan  
 "S/he does not love Joan" (Escandell-Vidal 2007: 29-30)

Therefore, the distribution of DOM in Balearic Catalan complies with our hypothesis that DOM with left-dislocated direct objects has a topic-shift or topic promotion function. The fact that DOM with left dislocated direct objects also encodes contrastive topics is not unexpected: by marking contrast against an active topic, DOM serves to re-activate an old one.

To sum up, DOM is compulsory with personal pronouns and some other NP classes in both Standard Peninsular and Balearic Catalan. DOM appears again with other NP classes provided that they are dislocated: this fact, once again, suggests that topicality is the main triggering parameter, along with animacy (Peninsular Catalan) and definiteness (Balearic Catalan). The situation found in Catalan suggests that the grammaticalisation of the construction has caused its narrowing to topical objects. This narrowing, however, is to be considered as a result of the standardisation of Catalan: since DOM was felt like a Castilian feature, its usage was strongly discouraged in the standardised language. As a matter of fact, DOM was much widely employed in older stages of Catalan. Meier (1945) provides quite a few examples of DOM in Old Catalan with proper names (8.34b), or human nouns (8.34a):

- (8.34) (a) *E per ço al bon àngel hom lo*  
 and for this DOM.the good angel IMPRS 3SG.M.OBJ  
*deu saludar tots jorns dues vegades, a*  
 must.PRS.3SG greet.INF every.PL days two times at  
*matí e vespre*  
 morning and evening

"And therefore, one should greet the good angel twice a day, in the morning and in the evening"

- (b) *E tots los apostols amaven tant a Jesuchrist*  
 and all the apostles love.IPFV.3PL much DOM Jesus  
*que no podien alçar los huylls al cel*  
 that NEG can.IPFV.3PL lift.INF the eyes to.the sky

"And all the apostles loved Jesus so much that they could not lift their eyes to heaven" (14th century; Meier 1945: 242)

As is clear from these examples, DOM was more widespread in Old Catalan. As a consequence of the standardisation process, thus, Catalan has retained DOM only in restricted contexts, regardless of the actual topicality of the overtly coded direct object. Still, spoken Catalan has grammaticalised the tendency for DOM to appear in high topical positions (a tendency already observable in 14th century Catalan, see example (8.34a), in which the direct object is dislocated), in that virtually every direct object, if dislocated, can be overtly coded.

Modern European Portuguese displays DOM in very limited contexts. Basically, it is obligatory when the direct object is the relative pronoun *quem*, as in (8.35a), a personal pronoun doubled by a clitic (8.35b), and when the direct object is topicalised or dislocated and resumed by a clitic within the clause, as in (8.35c, d) (Mira *et al.* 2003):

- (8.35) (a) *Vi o velhote \*(a) quem o Luís*  
 see.PST.1SG the old.man DOM REL.SG the Luís  
*ajudou*  
 help.PST.3SG  
 "I saw the man whom Luis helped"
- (b) *Vi-os \*(a) eles à saída do cinema*  
 see.PST.1SG-CLIT.3PL.OBJ DOM 3PL at exit of.the cinema  
 "I saw them at the exit of the cinema"
- (c) *Ao João, vejo sempre nos dias de eleições*  
 DOM.the João, see.PRS.1SG always in.the days of elections  
 "Joao, I always see in the days of elections"
- (d) *Ao João, vejo-o sempre nos dias*  
 DOM.the João, see.PRS.1SG-CLIT.3SG.OBJ always in.the days  
*de eleições*  
 of elections  
 "Joao, I always see him in the days of elections" (Mira *et al.* 2003:  
 286)

In Brazilian Portuguese, DOM is used with nouns referring to God (a usage attested in European Portuguese too) and with dislocated object pronouns, as in (8.36a, b), with or without the resumptive clitic within the clause (Flavia Teixeira, personal communication):

- (8.36) (a) *A mim não incomoda*  
 DOM 1SG NEG bother.PRS.3SG  
 "It doesn't bother me"
- (b) *A ele não levavam na onda. Era sabido*  
 DOM 3SG.M NEG take.IPFV.3PL the wave. be.PST.3SG clever  
*demais*  
 too  
 "They didn't fool him. He was too clever" (Thomas 1969: 256)

In Portuguese too, the usage of DOM is still closely related to the topicality of the direct object, as demonstrated by the fact that dislocated pronouns must be obligatorily overtly coded (8.37b), while post-verbal pronouns do not normally receive DOM (8.37a):

- (8.37) (a) *Viste ele?*  
 see.2SG.PST 3SG.M  
 "Did you see him?"

- (b) *A ele viste-o?*  
 DOM 3SG.M see.2SG.PST-CLIT.3SG  
 "Did you see him?" (Jose Florentino, personal communication)

Portuguese seems to have narrowed the usage of DOM to topical direct objects only, similarly to Catalan, as shown by the obligatory occurrence of DOM with dislocated direct objects as opposed to non-dislocated ones.

### 8.4.2 French varieties

Along with Standard and Northern Italian and Gallo-Italian dialects, French is considered one of the Romance languages where DOM does not appear at all. As a matter of fact, DOM is quite common with dislocated pronominal direct objects in Spoken French<sup>7</sup>, and has been extended to some lexical direct objects, such as kinship terms, in some regional varieties.

The presence of DOM in Southwestern (Toulouse, Bordeaux) varieties of French has been long considered as a result of influence of Gascon, which displays DOM (see Joly 1971 on Bearnese). Although an influence of Gascon may well be possible, it appears that DOM in Southwestern varieties is much more restricted in its distribution. While DOM in Gascon is used with human and definite direct objects (Rohlf's 1971), be they topical (dislocated) or focal (post-verbal), all the examples I have collected clearly show that DOM in the French varieties spoken in the same area is restricted to topical direct objects. The direct objects must be left- or right-dislocated, and doubled by a clitic, as shown by (8.38). Interestingly, DOM sporadically extends to human common nouns, as in (8.38e):

- (8.38) (a) *à moi, personne ne me veut*  
 DOM 1SG nobody NEG CLIT.1SG want.PRS.3SG  
 "Nobody wants me"
- (b) *à lui, on ne l'a voulu*  
 DOM 3SG.M IMPRS NEG 3SG.M.OBJ-AUX.1SG.PRS NEG  
 want.PTCP.PST  
 "As for him, they didn't want him"
- (c) *il faut l'aider, à elle*  
 3SG.SUBJ must.PRS.3SG 3SG.(F).OBJ-help.INF DOM 3SG.F  
 "As for her, we should help her"

<sup>7</sup>I wish to thank Vincent Gerbe, Solène Mantione, and Dominique Fazio for providing me with many of the examples discussed in this section, as well as Elise Steenackers for the examples from Brussels French. If not otherwise indicated, the examples are from my own notes.



- (d) *à ton père, je l'ai vu*  
 DOM your father 1SG.SUBJ 3SG.(M).OBJ-AUX.1SG  
 see.PTCP.PST  
 "As for your father, I saw him"
- (e) *Il le va blesser, à cet enfant*  
 3SG.SUBJ 3SG.M.OBJ AUX.3SG.PRS injure.INF DOM this  
 child  
 "He is going to injure this child" (Rolfhs 1971: 68)

As I said, an influence of Gascon on these usage cannot be excluded, even though the distribution of DOM in the Southwestern French varieties is narrower than in Gascon. By contrast, the presence of DOM in Brussels French, as well as in Spoken French, must be considered an independent development. Unlike Southwestern French, DOM in Brussels French and Spoken French is used only with dislocated personal pronouns resumed by a clitic within the clause, as in (8.39):

- (8.39) (a) *à lui, on ne l'attendait pas*  
 DOM 3SG.M IMPRS NEG 3SG.(M).OBJ\_wait.IPFV.3SG NEG  
 "We would not wait for him"
- (b) *Je la suivais, à elle*  
 1SG.SUBJ 3SG.F.OBJ follow.IPFV.1SG DOM 3SG.F  
 "I followed her"
- (c) *Il nous regardait toujours, à nous*  
 3SG.SUBJ 1PL look.IPFV.3SG always DOM 1PL  
 "He always used to look at us" (Hills 1920: 220)
- (d) *à moi, il ne m'attrapera pas*  
 DOM 1SG 3SG.SUBJ NEG CLIT.1SG catch.FUT.3SG NEG  
 "He won't catch me"
- (e) *à moi, personne ne me plaint!*  
 DOM 1SG nobody NEG CLIT.1SG pity.PRS.3SG  
 "Nobody feels sorry for me!"
- (f) *et je l'ai vu, à lui aussi, là à côté de moi...*  
 and 1SG 3G.(M).OBJ-AUX.1SG see.PTCP.PST DOM 3SG.M  
 too, there at side of 1SG

"And I saw him too, there, next to me"

(<http://tinyurl.com/5ukp4dg>)

(g) *à moi, ça me touche beaucoup*  
 DOM 1SG, this CLIT.1SG touch.PRS.3SG a\_lot

"As for me, I'm very touched by that"

As is clear from the examples in (8.39), the presence of DOM is triggered by the topical, dislocated, position of the direct object. Direct objects in their canonical, post-verbal (i.e. focal) position, cannot be overtly coded, as in Standard and Northern Italian and Catalan. A similar situation is attested in the French variety of Switzerland and in some *patois* of the French-speaking area of Switzerland (Valais, Genève), where DOM is restricted to dislocated personal pronouns, as in (8.40):

(8.40) (a) *no t'in yu, a te*  
 1PL CLIT.2SG-AUS.1PL see.PTCP.PST DOM 2SG  
 "We have seen you" (Valais)

(b) *a te t'an proeu yu*  
 DOM 2SG CLIT.2SG-AUS.1PL enough see.PTCP.PST  
 "We have seen enough of you" (Valais; Rohlf's 1971: 71)

(c) *il m'a vu à moi*  
 3SG.SUBJ CLIT.1SG-AUX.3SG see.PTCP.PST DOM 1SG  
 "He saw me" (Fribourg, Switzerland)

(d) *à nous, nous n'avons jamais fait ceci*  
 DOM 1PL 1PL NEG-AUX.1PL never do.PTCP.PST this  
 "We ourselves, (as for us) we have never done this!" (Fribourg;  
 Rohlf's 1971: 71)

Example (8.40d) proves particularly interesting, as the NP introduced by *à* is not a direct object, but rather an agent. This seemingly anomalous usage, which is found also in some Italian dialects (see below) demonstrates that speakers employ the object marker to signal the highly topical nature of the overtly coded NP.

Summing up, French varieties and Spoken French do display DOM, contrary to what has been claimed in the previous literature on the topic. Interestingly, the conditions of appearance of DOM seem to be the same as in Northern and Standard Italian and Gallo-Italian dialects. Indeed, in order for DOM to show up, the **pronominal** direct object must be **dislocated**. Unlike Northern and Standard Italian, though, DOM in French (as well as in Catalan) is also found on right-dislocated direct objects, a fact that deserves further investigation.

In the next section I will discuss the extent to which topicality still influences DOM in other Romance languages where DOM does not depend upon information structure any longer, such as Spanish. Furthermore, based on historical data, I will show that topicality is likely to be the primary triggering context in which DOM arose.

## 8.5 A diachronic look: DOM and topicality in Spanish and Sicilian

In the previous section I have argued that DOM is present at an incipient stage even in those Romance languages where DOM is usually held to be absent, like French and Italian. I further argued that, in these languages, DOM is brought about by information-structural factors, namely topicality. This is demonstrated by the restriction of DOM to dislocated direct objects. The situation found in Catalan is very interesting as the marking of focal objects is only restricted to pronouns and some other minor NP classes, whereas the marking of other NP classes is only allowed when they are dislocated.

In this section, I will briefly discuss the role of topicality in Romance languages where DOM is no longer primarily driven by information structure, like Spanish and Sicilian. I will also deal with the role of topicality in the diachronic development of DOM in these languages.

### 8.5.1 Sicilian

Sicilian nicely exemplifies the extension of DOM to every direct object, regardless of its actual topicality. DOM in Modern Sicilian is obligatory with pronouns, proper names, singular kinship terms, as in (8.41a, b, c), the negative quantifier *nuddru* "nobody" and the indefinite pronoun *unu* "one" (see Iemmolo 2010a for more examples). The marker is optional with plural kinship terms (8.41d) and human common nouns. DOM is not allowed with (in)animate and human indefinite non-specific nouns:

- (8.41) (a) *Chiddru chiama*            A    MMIA  
           that        call.PRS.3SG    DOM 1SG  
           "That guy's calling me"
- (b) *Canuscivu a Luvici*  
           know.IPFV.1SG    DOM Luigi  
           "I met Luigi"

- (c) *Arrubbaru a so cuscinu*  
 kidnap.PFV.3PL DOM POSS.3SG cousin  
 "His cousin was kidnapped"
- (d) *Arrubbaru (a) i so cuscini*  
 kidnap.PFV.3PL DOM the POSS.3SG cousins  
 "His cousins were kidnapped" (Iemmolo 2010a: 343-344)

In the contexts listed above, DOM is obligatory, even when the direct object is focal, as shown by the presence of DOM in (8.41a), where the direct object bears prosodic stress being in focus. Indeed, object pronouns cannot be used without the preposition.

Within the optional contexts (i.e. plural kinship terms and human common nouns), DOM becomes compulsory when the direct object is dislocated, as shown by examples (8.42a) as opposed to (8.42b). Dislocated direct objects are indubitably topical, as demonstrated by the fact that they are resumed by the clitic within the clause. As a matter of fact, a focal direct object can be preposed in Sicilian (Cruschina 2006), but it does not receive DOM, nor is it resumed by the clitic, as in (8.42c):

- (8.42) (a) *Arrubbaru (a) i so cuscini aieri*  
 kidnap.PFV.3PL DOM the POSS.3PL cousins yesterday  
 "His cousins were kidnapped yesterday"
- (b) *\*(A) i so cuscini l'arrubbarru*  
 DOM the POSS.3PL cousins 3PL.M.OBJ\_kidnap.PFV.3PL  
*aieri*  
 yesterday  
 "His cousins were kidnapped yesterday"
- (c) *\*A I SO CUSCINI arrubbarru aieri*  
 DOM the POSS.3PL cousins kidnap.PFV.3PL yesterday  
 "His cousins were kidnapped yesterday"

However, the overt coding of (left)-dislocated direct objects is restricted to the NP classes that receive (obligatory or optional) DOM in non-dislocated position. This shows that topicality is, at best, only a secondary factor in determining DOM. Thus DOM in Modern Sicilian is triggered mainly by humanness along with definiteness/specificity, while topicality plays a role only with the NPs which are optionally overtly coded (see below).

The situation in Old Sicilian was quite different. As shown in Iemmolo (2009, 2010b), based on a statistical analysis of DOM on a corpus of six 14th century texts, the most relevant parameter in Old Sicilian was topicality (detected through

dislocation). Indeed, neither humanness nor definiteness alone can motivate the use of DOM. Indeed, DOM was not compulsory with pronouns and proper names, and there are many occurrences in which inanimate direct objects are overtly coded:

- (8.43) (a) *Quilli ki volinu ke la republica*  
 those who want.PRS.3SG that the republic  
*sia salva secutinu mi*  
 be.SUBJ.PRS.3SG safe follow.3PL.PRS.SBJV 1SG  
 "Those who want to save the Republic should follow me"
- (b) *E Quintu Catulu, astutatu e aucisu*  
 and Quintus Catulus, extinguish.PTCP.PST and kill.PTCP.PST  
*ca appi Marcu Lepidu*  
 that have.PFV.3SG Marcus Lepidus  
 "And Quintus Catulus, after killing Marcus Lepidus..."
- (c) *Cusì suctirraru a killu corpu*  
 thus bury.PFV.3PL DOM that body/corpse  
 "Thus they buried that corpse" (Iemmolo 2009)

The strongest statistical correlation in Old Sicilian is between DOM and dislocation. As showed by Iemmolo (2010b), dislocated direct objects are by far the most overtly-coded NP class. Interestingly, nearly half of the occurrences of overtly-coded direct objects in dislocations are pronouns (8.44a, b), a fact that patterns with the distributions found in Italian, French, Catalan, etc.:

- (8.44) (a) *Et ad issu medemmi tuctu lu Senatu lu*  
 and DOM 3SG.M really all the Senate 3SG.M.OBJ  
*acumpagnau intra lu Capitoliu.*  
 accompany.PFV.3SG into the Capitol  
 "And as for him, all the Senators accompanied him into the Capitol"
- (b) *Et a cti li segreti cammari di li nostri dei*  
 and DOM you the unknown rooms of the our gods  
*ti aspectanu*  
 CLIT.2SG wait.PRS.3PL  
 "And the unknown rooms of our gods are waiting for you" (Iemmolo 2010a: 257)

Sicilian well exemplifies the extension of DOM to direct objects which have topic-worthiness features. While DOM in Old Sicilian was found with topical direct objects, Modern Sicilian has generalised DOM to a subset of direct objects

with features typical of topics, like humanness and definiteness, regardless of their information status.

As I have argued in detail in Iemmolo (2009), at some stage the information status of the direct object began losing importance as the grammaticalisation process went on. Thus topicality became only a secondary parameter, which can still be relevant in cases of optionality of the marker. As a matter of fact, topicality is often the factor that paves the way for the extension of DOM to further NP classes. That is, the extension of DOM to another NP class applies first to topical direct objects, and is later extended to non-topical ones. For instance, DOM in Old Sicilian was obligatory when the pronouns were dislocated and therefore topical. Later on, DOM was extended to every pronominal direct object. In Modern Sicilian, DOM is optional with plural kinship terms and human common nouns when *in situ* or focal. When topical, however, these direct objects obligatorily receive DOM. Thus one may surmise that the next step of the grammaticalisation of DOM will involve its spreading to plural kinship terms and human common nouns, irrespective of their topicality.

Similar distributions are found in other Romance languages with rich DOM systems, like Spanish, which will be discussed in the next section.

### 8.5.2 Spanish

DOM is one of the most studied phenomena of Spanish grammar (von Heusinger and Kaiser 2005, 2007, 2011; Laca 1995, 2006; Leonetti 2004, 2008; Pensado 1995, Torrego Salcedo 1999, among others), and it will be hardly possible to summarise the variety of proposals that have been put forward to account for its distribution and its governing conditions. DOM in Modern Peninsular Spanish is obligatory with definite human direct objects (Leonetti 2004, Torrego Salcedo 1999)<sup>8</sup>, while it is optional with indefinite specific and non-specific ones, as shown by the examples in (8.45):

- (8.45) (a) *Necesitan (a) un ayudante que sepa*  
 need.3PL.PRS DOM one assistant who know.SUBJ.PRS.3SG  
*inglés*  
 English  
 "They need an assistant who knows English"
- (b) *Está buscando a alguien*  
 be.PRS.3SG look.GER DOM someone  
 "S/he is looking for someone"

<sup>8</sup>DOM with inanimate direct objects is not uncommon (Company 2002, García García 2007, Weissenrieder 1991, among others) and suggests that DOM is extending downwards the animacy hierarchy. I will not go deeper into this issue here.

Similarly to Modern Sicilian, DOM becomes obligatory in pre-verbal position with direct objects for which DOM is optional when in post-verbal position, as in (8.46a) vs. (8.46b), and (8.46c) vs. (8.46d):

- (8.46) (a) *Ya conocía (a) muchos estudiantes*  
 already know.IPFV.1SG DOM many students  
 "I already knew many students"
- (b) \*(A) *muchos estudiantes, ya los conocía*  
 DOM many students already 3PL.OBJ know.IPFV.1SG  
 "Many students, I already knew "
- (c) *Habían incluido (a) dos catedráticos en la lista*  
 have.IPFV.3PL include.PTCP.PST DOM two professors in the list  
 "They included two professors in the list"
- (d) \*(A) *dos catedráticos, los habían incluido en la lista*  
 DOM two professors 3PL.OBJ have.IPFV.3PL include.PTCP.PST in the list  
 "Two professors they included in the list" (Leonetti 2004)

Thus DOM in Spanish still shows a secondary effect of topicality, similar to what we have observed in Sicilian, insofar as topicality renders obligatory the presence of the preposition with the NP classes that would be only optionally overtly coded. The chief role of topicality even in languages where DOM is primarily governed by semantic factors like animacy is indeed more evident when data from other varieties of Spanish are taken into consideration. In Porteño Spanish (spoken in Buenos Aires and the area near the Rio Plata), DOM has been extended to inanimate direct objects, provided that they are left-dislocated, as in (8.47) (Dumitrescu 1997, 1998). The use of DOM is thus triggered by the topical, dislocated position of the direct object. If the direct object appears in post-verbal position, and is focal, DOM is impossible, as shown by (8.47b) as opposed to (8.47a):

- (8.47) (a) *A esta plaza la cuidan Aerolíneas Argentinas y usted*  
 DOM this square 3SG.F.OBJ care.PRS.3PL Aerolíneas Argentinas and you  
 "You and Aerolíneas Argentinas take care of this square"

- (b) *Aerolíneas Argentinas y usted cuidan (\*a) esta plaza*
- (c) *A estas flores vulgarmente las*  
 DOM these flowers vulgarly 3PL.F.OBJ  
*llaman/denominan margaritas*  
 call.PRS.3PL/name.PRS.3PL daisies  
 "These flowers, they commonly call them daisies"
- (d) *A los libros los envolvió en papel madera*  
 DOM the books 3PL.OBJ wrap.PST.3SG in paper wood  
 "The books, she wrapped them in paper wood" (Dumistrescu 1997)

Similar developments are attested in other Latin-American varieties of Spanish, where the extension of DOM to inanimate direct objects seems to be strongly determined by topicality, as observed by Company (2002) and Laca (2006).

The role of topicality seems to hold in diachrony as well. As shown by Melis (1995), Pensado (1985), and Laca (2006), DOM in Old Spanish was primarily governed by the topicality of the direct object rather than animacy or definiteness. According to Pensado (1985), DOM in Spanish indeed arose in left-dislocations of personal pronouns to overtly code the direct object as a topic<sup>9</sup>.

The examples in (8.48) from the *Cantar de mio Cid* nicely illustrate the role of topicality. When the direct object is post-verbal, there is no DOM, as in (8.48a, c). By contrast, when the direct object is pre-verbal, as in (8.48b, d), DOM shows up as expected:

- (8.48) (a) *Ca yo case sus fijas con*  
 that 1SG.SUBJ marry.PST.1SG POSS.3.PL daughters with  
*yfantes de Carrion*  
 princes of Carrion  
 "That I married off his daughters to the Princes of Carrion" (*Cantar de mio Cid*, 2956; 13th century)
- (b) *Que a mis fijas bien las*  
 that DOM POSS.1.PL daughters good 3PL.F.OBJ  
*casare yo*  
 marry.FUT.1SG 1SG.SUBJ  
 "That my daughters, I will marry them well" (2834)

<sup>9</sup>In Pensado's (1985) view, the use of the preposition *a* is also due to the fact that Spanish lost the Latin distinction between accusative and dative in the pronominal paradigm. The origin of DOM in the confusion between dative and accusative in the pronominal paradigm, however, is not borne out by comparative Romance data. Indeed, there are languages with DOM where this neutralisation did not take place, as Rumanian or Sardinian, and languages where this neutralisation occurred but no DOM is detectable, as Friulian and some Rhaeto-Romance varieties (see Haiman and Benincà 1992).



(c) *En braços tenedes mis fijas tan blancas*  
 in arms hold.PRS.2PL POSS.1.PL daughters so white  
*commo el sol*  
 as the sun

"In your arms you hold my daughters as white as the sun" (2333)

(d) *A la sus fijas en braços las*  
 DOM the POSS.3.PL daughters in arms 3PL.F.OBJ  
*prendia*  
 take.IPFV.3SG

"His daughters he took in his arms" (275; examples from Melis 1995)

The synchronic and diachronic data presented thus far support the hypothesis that topicality, expressed through dislocation, was the origin of the construction in Spanish. However, due to grammaticalisation, DOM has been extended to human direct objects, as long as they are topical. Thus the primary trigger for DOM has become the animacy of the object. Topicality, however, still exerts a certain influence on the appearance of DOM in optional contexts, as we have discussed above. That is, topicality is the factor that paves the way for further extensions of DOM, even though it no longer takes priority over the other parameters. Even in Romanian, where DOM, expressed by the preposition *pe*, is primarily regulated by specificity and animacy, topicality still influences the appearance of DOM. Direct objects introduced by *pe* are indeed more likely to be taken up and commented upon in the subsequent discourse than uncoded ones, i.e. overtly coded direct objects show a higher degree of referential persistence (Chiriacescu and von Heusinger 2011).

## 8.6 Conclusion

We have seen that topicality turns out to be relevant in all the Romance languages I have examined thus far. The topicality hypothesis is further supported by the fact that, already in Classical Latin, the preposition *ad* (whose basic meaning, like its Romance descendants, revolves around motion towards a place), could be used to introduce an *as-for* topic, as in (8.49):

(8.49) (a) *Ad Dolabellam, ut scribis, ita*  
 ALL Dolabella.ACC.SG.F as write:PRS.2SG so  
*puto faciendum*  
 believe.PRS.1SG do.GER.ACC.SG.N

"As for Dolabella, as you write, I think that we should act in this way" (Cic. Att. 13, 10, 2; from Havers 1925)

(b) *Ad ea autem, quae scribis*  
 ALL DEM.ACC.PL.N instead REL.ACC.PL.N write.PRS.2SG  
*de testamento, videbis quid et quomodo*  
 about will.ABL.SG see:FUT.2SG what.ACC and how  
 "With regard to what you say about the will, please consider what should be done and how" (Cic. Att. 11, 21, 1; from Havers 1925)

(c) *Ita ad Capuam res*  
 thus ALL Capua.ACC.SG thing.NOM.PL  
*compositae consilio ab*  
 settle.PTCP.PFV.NOM.PL.F deliberation.DAT.SG from  
*omni parte laudabili*  
 every.ABL.SG side.ABL.SG commendable.DAT.SG  
 "In this way matters regarding Capua were settled with a plan worthy of commendation in every respect" (Tit. Liv. *Ab Urbe Condita* 26, 16, 11)

This usage was probably typical of spoken language, and is the likely starting point for the more extensive use of the preposition in Late Latin as a topic marker, subsequently as a recipient/beneficiary marker, and finally as a differential object marker in Romance languages (notice that these meanings, except for the differential object marker, coexist both in Latin and in Romance languages). The topic marking function was probably inferred from the directional meaning of the preposition, with a meaning like "(turning our attention) to/toward this" (see Ernout and Meillet 1985: 8). The following path can thus be postulated:

- locative, allative > (topic) > dative > (differential) direct object marker

This analysis provides further evidence for the idea that topicality is the property that links all these meanings. I am not claiming here that *ad* had a grammaticalised topic-marking function in Classical Latin. Nevertheless, based on cross-linguistic evidence, it seems to be likely that the usage of *ad* with a semantics like "as for, as to" was one of the factors that paved the way for the extension of this preposition to the dative, a role usually occupied by human, definite referents, and subsequently to topical direct objects. All these usages, except for the differential object marking function of *ad*, are already attested in Classical Latin. It should be noted that the dative stage is not obligatory, as we have seen in Chapter 6. For instance, it is not attested in Kanuri, Galo, as well as in other Tibeto-Burman or Indo-European languages, like Persian or Classical Armenian. Crucially, in all these languages, the dative was encoded in a different way from the accusative. Moreover, the triggering role of the topicality parameter is further documented by the (apparently) unusual usage found in some French varieties and *patois* (see

above, section 8.4.2) as well as in some Italian dialects. For example, in Campidanese Sardinian, the topicalising function of the preposition *a* has been extended to a prototypical topic function, that is, subject of an active clause (Putzu 2008: 412, quoting Blasco Ferrer 1984: 84). This is illustrated in (8.50):

- (8.50) *A chi arriidi urtimu, arriidi mellus*  
 DOM who laughs last laughs best  
 "He who laughs the last, laughs best" (Putzu 2008: 412)

As I have widely argued in the previous chapters, the topic-marking function of the differential object marker is cross-linguistically widespread. For example, the original meanings of Persian *-râ* and Hindi *-ko* object markers were "goal", "with respect to". In accordance with the hypothesis just presented, these adpositions further developed into dative and accusative markers (cf. Lehmann 1995: 98, Lazard 2001: 875, among others). The same process is attested in other Iranian languages, such as Shugni and Parâči, where the current object markers probably once meant *as for*" (Lazard 2001: 875), as well as in Armenian (Meillet 1903) and in Austronesian (Maori, Puyuma).

## Chapter 9

### Conclusion

As stated in the introduction to this study, the main aim of this research was to investigate the factors underlying the encoding of animate and definite direct objects from a typological perspective. In Chapters 1–2, the theoretical foundations and the methodological premises of the approach adopted in this study were outlined. In Chapter 3 I have presented the problems displayed by previous approaches and discussed possible solutions to these problems. I have proposed, following Croft (1988) and Siewierska (1997), that we should abandon the idea of a single characterisation of direct objects. I have shown that the grammatical behaviour of direct objects is sensitive to distinct functional and morpho-syntactic factors. In particular, I have argued that direct objects pattern differently with regard to case marking as opposed to indexation. Although both case marking and indexation tend to appear with topical direct objects, yet they have different functional motivations. Overt coding on topical direct objects is indeed motivated by the unexpectedness of highly topical/topic-worthy lexical direct objects, because of the low frequency of such NPs in direct object position. By contrast, indexation is correlated with a high degree of topicality of the referent it indexes. In other words, overt coding serves to signal that a lexical NP fulfilling the role of direct object is highly topical, while indexation is naturally associated with topical NPs and signals the discourse prominence of such direct objects.

As I have widely discussed in Chapter 3, the need for overt coding of topical direct objects is not due to the fact that a topical direct object is cognitively more marked per se. Rather, I have argued that it is the lower frequency of topical/topic-worthy referents in direct object position that explains why these direct objects are less easy to process. In addition, we have seen that language processing is based on the expectations and predictability of the language user. Thus, the bias towards predictability (Hume 2004: 3) envisages that grammatical relations in a clause will be assigned based on the expectations (built up through frequency) of the language user. Since the most frequent situation seems to be one in which

agents are human, definite and topical and direct objects are less human, definite and/topical than agents, the reversal of this situation would be likely to be harder to process.

The characterisation of DOM and DOI as two constructions differently sensitive to the referential and information-structural properties of the direct object allows us to solve and account for the seemingly paradoxical behaviour of direct objects with respect to case marking and indexation. Thus, DOM is governed by the unexpectedness of the properties of the direct object referent, whilst DOI is associated with the signalling of high salience or prominence. In turn, this functional differentiation is mirrored by the distinct discourse functions covered by the two constructions. The cross-linguistic analysis of DOM and DOI carried out in Chapters 5-7 has indeed shown that, although DOM and DOI are both determined by topicality, yet they differ as to their topic-marking functions. I have shown that DOM is overwhelmingly associated with devices for signalling topic discontinuities, such as topic-shift and topic-promotion structures. By contrast, DOI is consistently associated with highly continuous topics.

I have shown that, cross-linguistically, DOM is recurrently associated with two major construction types, namely topicalisations and dislocations, or with positions reserved to topical referents, such as the pre-verbal position in languages where DOM originates from former serial verb constructions. By contrast, DOI is associated with highly accessible referents, which indeed are normally referred to by zero-anaphora and analogous continuity markers. Hence, DOM serves to introduce or shift a topic, while DOI serves to encode the direct object as a normal topic expression (see Chapter 4) to maintain topic continuity throughout discourse.

The idea that topicality is the major factor underlying DOM and DOI systems is supported by diachronic evidence as well. In Chapter 6-8 I have shown that, in addition to being synchronically governed by topicality, cross-linguistically object markers are part of polysemy patterns involving topicality- or topic-related functions, such as frame-setting expressions (i.e. spatio-temporal expressions), as well as proper topic or conditional markers. I have also argued that, even if the connection between object markers and topic-related functions is no longer transparent synchronically, such a connection can still be reconstructed based on diachronic evidence, where available.

Further proof for the primacy of topicality over semantic parameters like animacy and/or definiteness comes from the analysis of systems synchronically based on animacy advanced in Chapter 5, where I showed that DOM and DOI systems that appear to be based on animacy are a result of the grammaticalisation of earlier topicality-based systems. Since human referents are the most topical, over time this led to the conventionalisation of DOM and DOI to human direct objects, irrespective of their information status. Thus, DOM and DOI grammaticalised with human direct objects following the animacy hierarchy discussed in Chapter 4. In

addition to this, I have also explored the role of topicality in the so-called optional contexts. As I said above, in many languages, mostly those examined in Chapter 5, DOM and DOI are primarily dependent upon the semantic features of the direct object, like animacy. In these languages, DOM and DOI are obligatory with some NPs (usually those higher on the animacy/definiteness hierarchies), and optional with less animate/definite NP classes. In my analysis, the appearance of DOM and DOI is in fact never optional. Rather, in these cases, topicality is the decisive factor in determining the usage of overt coding or indexation. That is, DOM or DOI show up only if the NP they overtly code and/or index is topical. As a consequence of increasingly high frequency, DOM and DOI further grammaticalise with that NP class and are no longer optional, being also reanalysed as neutral as to their information status. This process is attested in several unrelated languages (Indo-European, Dravidian, Nilo-Saharan, Tucanoan, Uralic, among others), as we have seen in Chapter 6.

A similar path of development was also examined in Chapter 8 devoted to Romance languages. Indeed, the data from Romance languages, in particular languages where DOM is at an incipient stage, like French varieties, Standard and Northern Italian, and Northern Italian dialects, clearly shows that DOM emerged as a strategy to code the topicality of the direct object overtly. In point of fact, DOM in Romance arose first with dislocated direct objects, and was later extended to direct objects showing features of topic-worthiness, as in Spanish or Sicilian. The extension, however, is still driven by topicality, since DOM first becomes obligatory with topical direct objects (i.e. dislocated), and is then extended to the post-verbal position, where semantic features like animacy or definiteness start becoming the decisive factors. In other languages, like Catalan and Portuguese, DOM underwent a process of narrowing, as shown by the optionality of DOM with post-verbal direct objects as opposed to dislocated ones. Thus, the rise and distribution of DOM in Romance comply with the model we have proposed in Chapter 3. On the whole, the approach advanced in this work offers an overarching explanation for DOM and DOI that differs in many respects from previous proposals made in the literature. As we have seen in Chapter 2, DOM has been analysed either as a distinguishing or an indexing strategy. Most work on DOM has indeed assumed that the primary motivation behind DOM is the need for distinguishing agents from direct objects when the latter are animate and/or definite. However, I have argued that this analysis is not supported by the data examined in this work, except for the cases of Kashmiri (Indo-Aryan) and Malayalam (Dravidian) (see sections 3.2 and 5.12 for further discussion). Likewise, I have shown that similar problems arise if one adopts the view that takes DOM and DOI as reflections of the high degree of affectedness of the direct object (see Chapter 7 for discussion on Mandarin Chinese).

Finally, my approach differs from the topicality-based model proposed by Dalrymple and Nikolaeva (2011). Although Dalrymple and Nikolaeva's (2011) model is based on the idea that topicality is the main parameter behind DOM and DOI, yet there are crucial differences in our understanding and motivations for the phenomena under investigation. In Dalrymple and Nikolaeva's view, DOM and DOI are means to encode the (primary or secondary) topicality of some direct objects and highlight the similarities between topical agents and topical direct objects. Nevertheless, I have suggested in Chapters 2–3 that lexical topical or topic-worthy direct objects tend to be rather infrequent in discourse and therefore unexpected. Furthermore, unlike Dalrymple and Nikolaeva, I have argued that DOM and DOI, the commonalities in the triggering factors notwithstanding, are distinct phenomena that cannot be lumped together, since they carry out distinct functions in discourse.

In sum, my approach is different from both previous approaches, in which DOM and DOI are either viewed as reflections of the cognitive and semantic marked status of the direct object (as in the distinguishing approach) or as means to indicate the high degree of affectedness of the overtly coded or indexed direct object (as in the indexing model). Likewise, my characterisation of DOM and DOI diverges from the topicality-based approach put forward by Dalrymple and Nikolaeva (2011).

Certainly, there is still much to do in the study of the rise and distribution of DOM and DOI systems. Therefore, I would like to mention a few areas for which further research is needed. There are at least four directions along which this research could be continued. First, the attested patterns may be verified on a probability language sample. The sample used in this study is indeed a variety sample, that was not designed to ascertain the statistical relevance of the grammatical phenomena under investigation. Hence, the findings of this study should not be taken as evidence for the cross-linguistic frequency and distribution of DOM and DOI. A probability sample might support the attested pattern of distribution of DOM and DOI and thus make the generalisations advanced in this study more powerful. In addition, new or different patterns may arise, which have not been found in this study.

The second direction in which this study may be continued is diachronic. I have pointed out at various points of this work a number of grammaticalisation processes, diachronic sources, and synchronic polysemy patterns for object markers, which have provided further support for the synchronic role of topicality in DOM and DOI systems. Further diachronic studies in the development of DOM and DOI systems would provide precious insights on the interaction between morpho-syntax, information structure and discourse organisation. From a synchronic perspective, further attention should be paid to the discourse functions of DOM and DOI in analysing the distribution of these phenomena both within

a single language and cross-linguistically. The increasing availability of corpora makes it possible to investigate the effects of the discourse context, as well as of the presence of case-marking on agents, on DOM and DOI.

Last but not least, the model proposed here could be assessed also through psycholinguistic and neurolinguistic experiments. Many studies on the processing of grammatical relations have dealt with the processing of arguments at the sentence level (see the overview in Bornkessel-Schlesewsky and Schlewsky 2009: 170-173), which have shown the importance of semantic factors like animacy in the processing of grammatical relations (Bornkessel-Schlesewsky and Schlewsky 2009a, 2009b and references therein). The inclusion of information-structural parameters such as topicality in the investigation of the processing of grammatical relations at the discourse level would indeed be crucial to account for the motivations underlying the distribution and the interaction of case marking vis-à-vis indexation and would also broaden our understanding of the functioning of these two phenomena.



# Appendix A

## Languages in the sample

The genetic affiliation of the languages in the sample is based on the website [www.ethnologue.com](http://www.ethnologue.com).

<b>Family</b>	<b>Genus</b>	<b>Language</b>
<b>Afro-Asiatic (19)</b>	<i>Cushitic (4)</i>	Alaaba
		Dullay
		Kemantney
		Ts'amakko
	<i>Egyptian-Coptic (1)</i>	Coptic (Sahidic)
		Amharic
	<i>Semitic (12)</i>	Aramaic (Biblical)
		Neo-Aramaic
		Arabic (Lebanese)
		Argobba
		Hebrew (Biblical)
		Hebrew (Modern)
		Maltese
Silti		
Tigre		
Tigrinya		
<i>Omotic (2)</i>	Zway	
	Haro	
	Sheko	
<b>Altaic (11)</b>	<i>Mongolic (4)</i>	Bao'an
		Buriat
		Mangghuer
		Mongolian (Khalkha)
	<i>Tungusic (2)</i>	Manchu
	Udihe	

	<i>Turkic (5)</i>	Altai Kirghiz Turkish Tuvan Uzbek
<b>Aracauan (1)</b>		Mapudungun
<b>Arawakan (1)</b>		Paumari
<b>Australian (7)</b>	<i>Pama-Nyungan</i>	Arabana Dharumbal Gunya Nhanda Pitjantjatjara Wargamay Waga-Waga
<b>Austro-Asiatic (2)</b>	<i>Aslian (1)</i> <i>Munda (1)</i>	Semelai Santali
<b>Austronesian (12)</b>	<i>Barito (1)</i> <i>Meso-Philippine (2)</i>  <i>Oceanic (6)</i>  <i>Northwest Malayo-Polynesian (1)</i> <i>Palauan (1)</i> <i>Sulawesi (1)</i>	Malagasy Hiliganyon Tagalog Marquesan Nêlêmwa Tamabo Tawala Tinrin Tobati Begak-Ida'an Palauan Selayarese
<b>Barbacoan (1)</b>	<i>Pasto</i>	Awa Pit
<b>Border (2)</b>	<i>Border (1)</i> <i>Waris (1)</i>	Imonda Waris
<b>Creoles (4)</b>	<i>Portuguese-based (3)</i>  <i>Malay-based (1)</i>	Diu Indo-Portuguese Kristang South African Creole Manadonese
<b>Dravidian (5)</b>	<i>Southern-Dravidian</i>	Badaga Betta Kurumba Kannada Malayalam Tamil
<b>Hokan (2)</b>	<i>Pomoan</i>	Central Pomo Eastern Pomo

<b>Indo-European (16)</b>	<i>Armenian (3)</i>  <i>Germanic (1)</i> <i>Indo-Aryan (5)</i>   <i>Iranian (5)</i>   <i>Slavic (2)</i>	Classical Armenian Eastern Armenian Western Armenian Afrikaans Kashmiri Hindi Maithili Marathi Romani Balochi Persian Tajik Vafsi Wakhi (Wakhan, Hunza) Nashta Pomak
<b>Iroquoian (1)</b>	<i>Northern Iroquoian</i>	Mohawk
<b>Isolate (4)</b>		Kwaza Purepecha Takelma Zuni
<b>Maku (2)</b>		Hup Dâw
<b>Muskogean (2)</b>	<i>Western (1)</i> <i>Eastern (1)</i>	Choctaw Koasati
<b>Na-Dene (2)</b>	<i>Northern Athabaskan</i>	Babine Witsuwit'en Slave
<b>Niger-Congo (16)</b>	<i>Atlantic-Congo (1)</i> <i>Bantoid (9)</i>          <i>Dogon (1)</i> <i>Kwa (3)</i>   <i>Idomoid (1)</i>	Noon Chichewa Kinande Makhwa Ngiemboon Sambaa Setawana Swahili Tswana Zulu Dogon (Donno-So) Akan Baule Ga Idoma

	<i>Nupoid (1)</i>	Nupe
<b>Nilo-Saharan (8)</b>	<i>Fur (1)</i> <i>Maban (1)</i> <i>Nilotic (1)</i> <i>Nubian (2)</i>  <i>Taman (1)</i> <i>Western Saharan (2)</i>	Fur Maba Lango Dongolese Nubian Uncunwee Tama Kanuri Tubu
<b>Oto-Manguean (3)</b>	<i>Mixtecan (1)</i> <i>Zapotecan (2)</i>	Copala Trique Yaitepec Chatino Zenzontepec Chatino
<b>Penutian (3)</b>	<i>Klamath-Modoc (1)</i> <i>Sahaptin (2)</i>	Klamath Nez Perce Yakima
<b>Sepik (4)</b>	<i>Ndu (1)</i> <i>Ramu(1)</i> <i>Tama (1)</i> <i>Yellow River (1)</i>	Manambu Awtuw Yessan Mayo Namia
<b>Sino-Tibetan (18)</b>	<i>Chinese (6)</i>     <i>Burmese-Lolo (3)</i>     <i>Bodic (7)</i>          <i>Kuki-Chin (1)</i> <i>Tani (1)</i>	Cantonese Gan Hakka Mandarin Southern Min Xiang Burmese Lahu Yongren Lolo Chantyal Chepang Newari (Dolakha) Kham Magar Manange Thulung Rai Meithei Galo
<b>Trans New Guinea (9)</b>	<i>Alor-Pantar (2)</i>  <i>Madang (2)</i>  <i>Mek (1)</i>	Abui Teiwa Anamuxra Tauya Una

	<i>Oksapmin (1)</i>	Oksapmin
	<i>Southeast Papuan (2)</i>	Barai Daga Dani
	<i>West Dani (1)</i>	
<b>Tucanoan (8)</b>	<i>Central (1)</i> <i>Eastern (4)</i>  <i>Western (3)</i>	Cubeo Barasano Desano Kotiria Tucano Koreguaje Retuarã Siona
<b>Tupi (1)</b>	<i>Tupi-Guaraní</i>	Guaraní
<b>Uralic (10)</b>	<i>Permic (1)</i> <i>Samoyedic (4)</i>  <i>Ugric (5)</i>	Komi Enets Nenets Nganasan Selkup Hungarian Mansi (Eastern) Mansi (Northern) Ostyak/Khanty (Eastern) Ostyak/Khanty (Northern)
<b>Yukaghir (1)</b>	<i>Southern</i>	Kolyma Yukaghir

Table A.1: Genetic affiliation of the languages in the sample

## Appendix B

### Main sources of information on the languages in the sample

Language	Source
Alaaba	Schneider-Blum 2007
Abui	Klamer and Kratochvíl 2006; Kratochvíl 2007
Afrikaans	den Besten 2000
Akan	Lord 1993
Altai	Skribnik 2001b
Amharic	Amberber 2005, Cohen 1936, Leslau 1995
Anamuxra	Ingram 2001
Arabana	Hercus 1994
Arabic (Lebanese)	Koutsoudas 1967
Aramaic (Biblical)	Khan 1984
Argobba	Appleyard 2004
Awa Pit	Curnow 1997
Awtuw	Feldman 1983
Babine Witsuwit'en	Gunlogson 2001
Badaga	Pilot-Rachoor 1994
Balochi	Korn 2008; Korn and Jahani 2009
Bao'an	Fried 2011
Barai	Olson 1981
Barasano	Jones and Jones 1991
Baule	Creissels ms.; Creissels and Kouadio 2010
Begak-Ida'an	Goudswaard 2005
Betta Kurumba	Coelho 2003
Buriat	Poppe 1960; Skribnik 2001b
Burmese	Jenny 2009; Thurgood 1978
Cantonese	Chappell 2006; to appear

Central Pomo	Mithun 1999
Chantyal	Noonan 1991
Chepang	Caughley 1982; Thompson 1990
Chichewa	Bresnan and Mchombo 1987; Mchombo 2004
Choctaw	Broadwell 2006
Classical Armenian	Meillet 1903; Thomson 1989
Copala Trique	Hollenbach 1992; López Cruz 2008
Coptic (Sahidic)	Engsheden 2008; Layton 2000
Cubeo	Morse and Maxwell 1999
Daga	Murane 1974
Dani	Bromley 1981
Dâw	Martins and Martins 1999
Desano	Miller 1999
Dharumbal	Terrill 2002
Diu Indo-Portuguese	Cardoso 2009
Dogon (Donno-So)	Plungian 1995
Dongolese Nubian	Dimmendaal 2010
Dullay	Tosco 1994, Amborn et al. 1980
Eastern Armenian	Dum-Tragut 2009; Megerdooomian 1999
Eastern Pomo	McLendon 1975, 1978
Enet	Körtvély 2005
Fur	Jakobi 1999; Waag 2009
Ga	Lord 1993
Galo	Post 2007
Gan	Chappell 2006; to appear
Guaraní	Shain and Tonhauser 2010
Gunya	Breen 1981
Hakka	Chappell 2006; to appear
Haro	Woldemariam 2004, 2009
Hebrew (Biblical)	Elwolde 1994; Waltke and O'Connor 1990
Hebrew (Modern)	Danon 2002; Ginert 1989
Hiliganyon	Spitz 2002
Hindi	Butt 1993; Mohanan 1994
Hungarian	Coppock and Wechsler to appear
Hup	Epps 2008; 2009
Idoma	Lord 1993
Imonda	Seiler 1985
Kannada	Lidz 2006
Kanuri	Hutchinson 1981; 1986
Kashmiri	Wali and Koul 1997
Kemantney	Leyew 2003

Kham	Watters 2002
Kinande	Baker 2003
Kirghiz	Imart 1981
Klamath	Rude 1988; Underriner 2002
Koasati	Kimball 1991
Kolyma Yukaghir	Maslova 2003b
Komi	Klumpp 2009, to appear
Koreguaje	Cook and Levinsohn 1985; Cook and Criswell 1993
Kotiria	Stenzel 2008
Kristang	Baxter 1988; 1994
Kwaza	van der Voort 2004
Lahu	Matisoff 1973
Lango	Noonan 1992
Maba	Weiss 2009
Magar	Grunow-Hårsta 2004
Maithili	Yadav 1996
Makhuwa	Stucky 1981; van der Wal 2009
Malagasy	Keenan 2008; Zribi-Hertz and Mbolatiavalona 1999
Malayalam	Asher and Kumari 1997
Maltese	Borg and Mifsud 2002
Manadonese	Wantalangi 1993
Manambu	Aikhenvald 2008
Manange	Hildebrandt 2004
Manchu	Gorelova 2002
Mandarin	Li and Thompson 1981; Li 2001
Mangghuer	Slater 2003
Mansi (Eastern)	Virtanen submitted
Mansi (Northern)	Skribnik 2001a
Mapudungun	Zúñiga 2010
Marathi	Pandharipande 1997
Margany	Breen 1981
Marquesan	Cablitz 2006
Meithei	Bhat and Ningomba 1997; Chelliah 1997, 2009
Mohawk	Baker 1996
Mongolian (Khalkha)	Guntsetseg 2008
Namia	Tupper 2009
Nashta	Adamou 2009
Nêlêmwa	Bril 2004
Nenets	Körtvély 2005
Neo-Aramaic	Coghill 2010, to appear
Newari (Dolakha)	Genetti 1997, 2007



Nez Perce	Rude 1982, 1986
Nganasan	Körtvély 2005
Ngiemboon	Watters 2003
Nhanda	Blevins 2004
Noon	Soukka 2000
Nupe	Güldemann 2007
Oksapmin	Loughnane 2009
Ostyak (Eastern)	Gulya 1966
Ostyak (Northern)	Nikolaeva 1999, 2001
Palauan	Nuger 2007, 2009; Woolford 2000
Paumarí	Chapman and Derbyshire 1991, Chapman 2008
Persian	Karimi 1990; Mahootian 1997
Pitjantjatjara	Bowe 1990
Pomak	Adamou 2009
Purepecha	Chamoreau 1999, 2000
Retuarã	Strom 1992
Romani	Matras 2002
Sambaa	Riedel 2009
Santali	Neukom 2000
Selayarese	Basri and Finer 1987; Finer 1997
Selkup	Körtvély 2005
Semelai	Kruspe 1999
Setawana	Demuth and Johnson 1989
Sheko	Hellenthal 2010
Silti	Appleyard 2004
Siona	Wheeler 1970
Slave	Rice 1989, 2003
South African Creole	den Besten 2000
Southern Min	Chappell 2006; to appear
Swahili	Marten and Kula ms.; Wald 1979, 1997
Tagalog	Himmelman 2005; Schachter 1972
Tajik	Windfuhr and Perry 2009
Takelma	Culy 2000
Tama	Dimmendaal 2009
Tamabo	Jauncey 2002
Tamil	Lehman 1993; Schiffman 1999
Tauya	MacDonald 1993
Tawala	Ezard 1991
Teiwa	Klamer 2010
Thulung Rai	Lahaussois 2002
Tigre	Jake 1980

Tigrinya	Kifle 2007; Kievit and Kievit 2009
Tinrin	Osumi 1995
Tobati	Donohue 2002
Ts'amakko	Savà 2005
Tswana	Demuth 1989
Tubu	Lukas 1953, König 2009
Tucano	Ramirez 1997; Zúñiga 2007
Turkish	Kornfilt 1997; von Heusinger and Kornfilt 2005
Tuvan	Harrison 2001
Udihe	Nikolaeva and Tolskaya 2001
Una	Louwerse 1988
Uncunwee	Jakobi 2009
Uzbek	Ismatullaev and Feldman 1995
Vafsi	Stilo 2010
Waga-Waga	Wurm 1976
Wakhi (Wakhan, Hunza)	Bashir 2009
Wargamay	Dixon 1981
Waris	Brown 1981, 1988
Western Armenian	Nilsenova 2002
Xiang	Chappell 2006; to appear
Yaitepec Chatino	Rasch 2002
Yakima	Jansen 2010
Yessan Mayo	Foreman 1974
Yongren Lolo	Gerner 2008
Zenzontepec Chatino	Carleton and Waksler 2000, 2002
Zulu	Buell 2006
Zuni	Nichols 1997
Zway	Appleyard 2004

Table B.1: Main bibliographical sources for the languages in the sample

# Appendix C

## List of the constructions analysed

Table C.1 provides an overview of the construction(s) found in each language, along with the main parameter(s) influencing the distribution of DOM and DOI. The following abbreviations are used:

**An** animacy

**Top** topicality

**Disl/Topic** dislocation/topicalisation or topical position

**Def** definiteness

**Aff** affectedness.

Language	Construction	An	Top	Disl/Topic	Def	Aff
Alaaba	DOI	+				
Altai	DOM		+	+		
Abui	DOI	+				+
Afrikaans	DOM	+				
Akan	DOM		+	+		
Amharic	DOM		+	+		
Anamuxra	DOM & DOI	+				
Arabana	DOM	+				
Arabic (Lebanese)	DOM & DOI	+				
Aramaic (Biblical)	DOM	+				
Argobba	DOM		+		+	
Awa Pit	DOM	+				
Awtuw	DOM	+				
Babine Witsuwit'en	DOI		+	+		

Badaga	DOM		+	+		
Balochi	DOM		+	+		
Bao'an	DOM		+	+		
Barai	DOI	+				
Barasano	DOM		+	+		
Baule	DOM		+	+		
Begak-Ida'an	DOM	+				
Betta Kurumba	DOM	+				
Buriat	DOM		+	+		
Burmese	DOM		+	+		
Cantonese	DOM		+	+		
Central Pomo	DOI	+				+
Chantyal	DOM		+	+		
Chepang	DOM		+	+		
Chichewa	DOI		+	+		
Choctaw	DOM		+	+		
Classical Armenian	DOM		+	+		
Copala Trique	DOM	+				
Coptic (Sahidic)	DOM		+	+		
Cubeo	DOM		+	+		
Daga	DOI	+				
Dani	DOI	+				
Dâw	DOM		+	+		
Desano	DOM		+	+		
Dharumbal	DOM	+				
Diu Indo-Portuguese	DOM	+				
Dogon (Donno-So)	DOM		+	+		
Dongolese Nubian	DOM		+			
Dullay	DOM		+	+		
Eastern Armenian	DOM		+	+		
Eastern Pomo	DOI	+				+
Enet	DOI		+			
Fur	DOM		+			
Ga	DOM		+			
Galo	DOM		+			
Gan	DOM		+			
Guaraní	DOM		+			
Gunya	DOM	+				
Hakka	DOM		+	+		
Haro	DOM		+			+
Hebrew (Biblical)	DOM		+	+		

Hebrew (Modern)	DOM				+	
Hiliganyon	DOM	+				
Hindi	DOM	+				
Hungarian	DOI				+	
Hup	DOM	+				
Idoma	DOM		+	+		
Imonda	DOM	+				
Kannada	DOM	+				
Kanuri	DOM		+	+		
Kashmiri	DOM	+				
Kemantney	DOM		+	+		
Kham	DOM		+			
Kinande	DOI		+	+		
Kirghiz	DOM		+	+		
Klamath	DOM		+	'		
Koasati	DOM		+	+		
Kolyma Yukaghir	DOM		+			
Komi	DOM		+	+		
Koreguaje	DOM		+	+		
Kotiria	DOM		+			
Kristang	DOM	+				
Kwaza	DOM	+				
Lahu	DOM		+			
Lango	DOI	+				
Maba	DOM		+	+		
Magar	DOM	+				
Maithili	DOM	+				
Makhuwa	DOI	+				
Malagasy	DOM				+	
Malayalam	DOM	+				
Maltese	DOM	+				
Manadonese	DOM	+				
Manambu	DOM		+	+		
Manange	DOM	+				
Manchu	DOM		+	+		
Mandarin	DOM		+	+		
Mangghuer	DOM		+	+		
Mansi (Eastern)	DOM & DOI		+	+		
Mansi (Northern)	DOI		+	+		
Mapudungun	DOI		+			
Marathi	DOM	+				

Margany	DOM	+				
Marquesan	DOM		+	+		
Meithei	DOM	+				
Mohawk	DOI	+				
Mongolian (Khalkha)	DOM		+	+		
Namia	DOM		+	+		
Nashta	DOM	+				
Nêlêmwa	DOI		+	+		
Nenets	DOI		+	+		
Neo-Aramaic	DOM & DOI		+			
Newari (Dolakha)	DOM		+			
Nez Perce	DOM		+			
Nganasan	DOI		+			
Ngiemboon	DOI		+			
Nhanda	DOM	+				
Noon	DOI	+				
Nupe	DOM		+	+		
Oksapmin	DOM	+				
Ostyak (Eastern)	DOI		+			
Ostyak (Northern)	DOI		+			
Palauan	DOM & DOI	+				+
Paumarí	DOM		+	+		
Persian	DOM		+	+		
Pitjantjatjara	DOM	+				
Pomak	DOM	+				
Purepecha	DOM		+	+		
Retuarã	DOM		+			
Romani	DOM		+			
Sambaa	DOI	+	+			
Santali	DOI	+				
Selayarese	DOI		+	+		+
Selkup	DOI		+			
Semelai	DOM		+			
Setawana	DOI		+	+		
Sheko	DOM		+	+		
Sil̥t̥i	DOM		+	+		+
Siona	DOM					
Slave	DOI		+	+		
South African Creole	DOM	+				
Southern Min	DOM		+	+		
Swahili	DOI	+	+	+		

Tagalog	DOM	+				
Tajik	DOM		+	+		
Takelma	DOI		+			
Tama	DOM		+			
Tamabo	DOI		+	+		
Tamil	DOM	+				
Tauya	DOI	+				
Tawala	DOI		+	+		
Teiwa	DOI	+				
Thulung Rai	DOM	+				
Tigre	DOM & DOI		+	+		
Tigrinya	DOM & DOI		+	+		
Tinrin	DOI		+	+		
Tobati	DOM		+	+		
Ts'amakko	DOM		+	+		
Tswana	DOI		+	+		
Tubu	DOM		+	+		
Tucano	DOM		+			
Turkish	DOM		+	+		
Tuvan	DOM		+			
Udihe	DOM		+	+		
Una	DOI	+				
Uncunwee	DOM		+			
Uzbek	DOM		+			
Vafsi	DOM		+			
Waga-Waga	DOM	+				
Wakhi (Wakhan, Hunza)	DOM		+			
Wargamay	DOM	+				
Waris	DOM	+				
Western Armenian	DOM		+	+		
Xiang	DOM		+	+		
Yaitepec Chatino	DOM	+				
Yakima	DOM		+			
Yessan Mayo	DOM	+				
Zenzontepec Chatino	DOM		+	+		
Zulu	DOI		+	+		
Zuni	DOM	+				
Zway	DOM		+		+	

Table C.1: Summary of the constructions found in the languages of the sample





# Appendix D

## Questionnaires

Below are the questionnaires used during my research on DOM in Northern Italian dialects (in Italian). Informants were asked to translate this questionnaire in their local dialect and judge translated sentences taken from other questionnaires.

### D.1 Questionnaire 1

In this questionnaire, various kinds of dislocated or topicalised direct objects (pronouns, proper nouns, kinship terms, common human nouns) are overtly coded.

1. A me, mi conoscono tutti
2. A lui, lo conoscono tutti
3. A lui, conosce, non a lei!
4. A me, m'annoia tutto questo
5. A te, ti secca tutto questo
6. A loro, gli secca tutto questo
7. A noi, ci hanno sempre invidiato
8. A noi, ci invidiano
9. Al piccolino, l'ha spaventato facendo così
10. A voi, vi ha spaventato
11. A lei, invece non l'ha spaventata affatto!
12. A me, m'ha offeso moltissimo il suo comportamento

13. A loro, li ha offesi moltissimo
14. A te, ti chiama sempre tutti i giorni
15. A Luigi, lo ha stupito tantissimo questo fatto
16. A papà, lo hai visto?
17. A te, t'ha seccato molto questa situazione?
18. A loro, li ha preoccupati tantissimo
19. A me, m'ha preoccupato tantissimo
20. Al ladro, non l'ha visto nessuno?
21. A lei, non l'ha vista nessuno?
22. A voi, non vi riguarda questa questione
23. A me, non mi convincono tutti questi ragionamenti
24. A loro, li ha fatti incavolare (incazzare) una cosa che non esiste
25. A lui, l'hanno ucciso con una violenza inaudita.

## D.2 Questionnaire 2

In this questionnaire, the overtly coded direct object consistently occupies the post-verbal position in the sentence. All these sentences were judged ungrammatical by speakers, who did not employ DOM in any case.

1. Conoscono tutti a me in questo posto
2. Mi conoscono tutti, a me
3. Conosce a lui, non a lei!
4. Annoia a me tutto questo!
5. Secca a me tutto questo
6. Hanno sempre invidiato a noi
7. Ha spaventato al piccolino facendo così
8. Ha spaventato a voi
9. Non ha spaventato affatto a lei

10. Il suo comportamento ha offeso moltissimo a me
11. Chiama tutti i giorni a me
12. Li ha offesi moltissimo, a loro, il suo comportamento
13. Questa cosa l'ha stupito tantissimo a Luigi.
14. Hai visto a papà?
15. Tutti questi ragionamenti non convincono a me
16. Nessuno ha visto al ladro?
17. Hanno ucciso a lui con una violenza inaudita
18. Questa questione non riguarda a voi
19. Questa situazione ha preoccupato molto a loro
20. Nessuno ha visto a lei?

## References

- Abondolo, Daniel M (ed.). 1998. *The Uralic languages*. Routledge Language Family Descriptions. London: Routledge.
- Adamou, Evangelia. 2009. Le marquage différentiel de l'objet en nashta et en pomaque. retour sur l'hypothèse du contact. *Bulletin de la Société de Linguistique de Paris* (104): 383–410.
- Aikhenvald, Alexandra Y. 1994. Grammatical relations in Tariana. *Nordic Journal of Linguistics* 17: 201–217.
- Aikhenvald, Alexandra Y. 2003. *A grammar of Tariana*. Cambridge: Cambridge University Press.
- Aikhenvald, Alexandra Y. 2008. *The Manambu Language of East Sepik, Papua New Guinea*. Cambridge: Oxford University Press.
- Aissen, Judith. 2003. Differential object marking: iconicity vs. economy. *Natural Language and Linguistic Theory* 21: 435–448.
- Allen, William S. 1951. A study in the analysis of Hindi sentence structure. *Acta Linguistica Hafniensia* 6: 68–86.
- Amberber, Mengistu. 2005. Differential subject marking in Amharic. In Amberber, Mengistu and Helen de Hoop (eds.), *Competition and variation in natural languages: the case for case*: 295–319. London: Elsevier.
- Amborn, Gunter Minker and Hans-Jürgen Sasse. 1980. *Das Dullay: Materialien zu einer ostkuschitischen Sprachgruppe*. Berlin: Dietrich Reimer Verlag.
- Ameka, Felix. 2010. Information packaging constructions in Kwa: micro-variation and typology. In Aboh, Enoch Oladé and James Essegbey (eds.), *Topics in Kwa syntax*, Studies in Natural Language and Linguistic Theory: 141–176. Dordrecht: Springer.
- Anderson, John. 1984. Objecthood. In Plank, Frans (ed.), *Objects. Towards a theory of grammatical relations*: 29–54. New York: Academic Press.
- Andorno, Cecilia. 2003. *Linguistica testuale*. Roma: Carocci.
- Appleyard, David. 2004. Some thoughts on the origin of the Amharic "object marker" [-ən]. In Böll, Verena, Denis Nosnitsin Thomas Rave Wolbert Smidt and Evgenia Sokolinskaia (eds.), *Studia Aethiopica. In honour of Siegbert Uhlig on the occasion of his 65th birthday*: 291–301. Wiesbaden: Harrassowitz.
- Arcodia, Giorgio Francesco and Giorgio Iemmolo. submitted. Differential object marking and identifiability of the referent: a study of Mandarin Chinese. *Linguistics*.
- Asbury, Anna. 2006. Finnish partitive case as a determiner suffix. In Gyuris, Beata (ed.), *Proceedings of the 1st Central European Student Conference in Linguistics*.
- Asher, Ronald E. and T. C. Kumari. 1997. *Malayalam*. London: Routledge.

- Badia Margarit, Antoni. 1994. *Gramàtica de la llengua catalana*. Barcelona: Proa.
- Baker, Mark. 1996. *The polysynthesis parameter*. Oxford Studies in Comparative Syntax. New York: Oxford University Press.
- Baker, Mark. 2003. Agreement, dislocation, and partial configurationality. In Carnie, Andrew, Heidi Harley and Mary Ann Willie (eds.), *Formal approaches to function in grammar. In honor of Eloise Jelinek*, *Linguistik Aktuell/Linguistics Today*: 107–132. Amsterdam-Philadelphia: John Benjamins.
- Baker, Mark. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.
- Bakker, Dik. 2010. Language sampling. In Song, Jae Jung (ed.), *The Oxford Handbook of Linguistic Typology*: 100–127. Oxford: Oxford University Press.
- Ball, Douglas. 2007. On ergativity and accusativity in Proto-Polynesian and Proto-Central Pacific. *Oceanic Linguistics* 46(1): 128–153.
- Barlow, Michael. 1992. *A situated theory of agreement*. New York: Garland.
- Bashir, Elena. 2009. Wakhi. In Windfuhr, Gernot (ed.), *The Iranian languages*: 825–862. London: Routledge.
- Basri, Hasan and Daniel L. Finer. 1987. The definiteness of trace. *Linguistic Inquiry* 18(1): 141–147.
- Baxter, Alan N. 1988. *A grammar of Kristang (Malacca Creole Portuguese)*. Canberra: Australian National University.
- Baxter, Alan N. 1994. Um importante sincretismo no português crioulo de malaca: a preposição multifuncional *ku*? In Cunha, Celine (ed.), *Miscelânea Celso Cunha*: 15–33. Rio de Janeiro: Melhoramentos.
- Beaudoin-Lietz, Christa, Derek Nurse and Sarah Rose. 2004. Pronominal object marking in Bantu. In Akinbiyi, Akinlabi and Oluseye Adesola (eds.), *Proceedings of the 4th World Congress of African Linguistics*, Cologne, 175–188. Rüdiger Köppe Verlag.
- Beavers, John. 2006. *Argument/Oblique Alternations and the Structure of Lexical Meaning*. Ph. D. thesis, Stanford University.
- Beavers, John. 2010. The structure of lexical meaning: Why semantics really matters. *Language* 86: 821–864.
- Beavers, John. to appear. On affectedness. *Natural Language and Linguistic Theory*.
- Belloro, Valeria A. 2007. *Spanish clitic doubling: A study of the syntax-pragmatics interface*. Ph. D. thesis, University of New York at Buffalo.
- Benincà, Paola. 1988. L'ordine degli elementi della frase e le costruzioni marcate. In Renzi, Lorenzo (ed.), *Grande grammatica italiana di consultazione*: 115–195. Il Mulino.
- Benincà, Paola. 2001. The position of topic and focus in the left periphery. In Cinque, Guglielmo and Giampaolo Salvi (eds.), *Current studies in Italian syntax*.

- Essays offered to Lorenzo Renzi*: 39–64. Amsterdam: Elsevier.
- Bentley, Mayrene E. 1994. *The syntactic effects of animacy in Bantu languages*. Ph. D. thesis, Indiana University.
- Berretta, Monica. 1989. Sulla presenza dell'accusativo preposizionale in italiano settentrionale: note tipologiche. *Vox Romanica* 13(1): 13–37.
- Berretta, Monica. 2002. *Temî e percorsi della linguistica. Scritti scelti*. Vercelli: Mercurio.
- Berry, Jeffrey J. 2006. Transitivity in Biblical Hebrew. In *Papers from the University of Utah Student Conference in Linguistics*, Salt Lake City.
- Bhat, Shankara, D.N. and M.S. Ningomba. 1997. *Manipuri Grammar*. Munich-Newcastle.
- Bhat, Shankara D. N. 1991. *Grammatical relations: the evidence against their necessity and universality*. London: Routledge.
- Bickel, Balthasar, Walter Bisang and Yogendra P. Yadava. 1999. Face vs empathy: the social foundation of Maithili verb agreement. *Linguistics* 37: 481–518.
- Blake, Barry J.. 1994. *Case*. Cambridge: Cambridge University Press.
- Blevins, Juliette. 2001. *Nhanda: an aboriginal language of Western Australia*. Honolulu: University of Hawaii Press.
- Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.
- Bokamba, Eyamba G. 1981. *Aspects of Bantu syntax*. Department of Linguistics: University of Illinois at Urbana-Champaign.
- Borg, Albert and Manwel Mifsud. 2002. Maltese object marking in a Mediterranean context. In Ramat, Paolo and Thomas Stolz (eds.), *Mediterranean languages: Papers from the MEDTYP workshop*, Diversitas Linguarum, 1: 33–46. Bochum: N. Brockmeyer.
- Börjars, Kersti and Nigel Vincent. 2008. Objects and OBJ. In Butt, Miriam and Tracy Holloway King (eds.), *Proceedings of LFG08*, Stanford, 150–168. CSLI Publications.
- Bornkessel-Schlesewsky, Ina and Matthias Schlewsky. 2009a. *Processing syntax and morphology: a neurocognitive perspective*. Oxford surveys in syntax and morphology. Oxford: Oxford University Press.
- Bornkessel-Schlesewsky, Ina and Matthias Schlewsky. 2009b. The role of prominence information in real time comprehension of transitive constructions: a cross-linguistic approach. *Language and Linguistic Compass* 3(1): 19–58.
- Bossong, Georg. 1985. *Differentielle objektmarkierung in den Neuiranischen Sprachen*. Tübingen: Narr.
- Bossong, Georg. 1991. Differential object marking in Romance and beyond. In Wanner, Dieter and Douglas A. Kibbee (eds.), *New analyses in Romance Linguistics: selected papers from the XVIII linguistic symposium on Romance languages*: 143–170. Amsterdam-Philadelphia: John Benjamins.

- Bossong, Georg. 1998. Le marquage différentiel de l'objet dans les langues d'Europe. In Feuillet, Jack (ed.), *Actance et valence dans les langues d'Europe*: 193–258. Berlin-New York: Mouton de Gruyter.
- Bowe, Heather. 1990. *Categories, Constituents, and Constituent Order in Pitjantjatjara, an Aboriginal Language of Australia*. London: Routledge.
- Breen, John Gayan. 1981. Margany and Gunya. In Dixon, R. M. W. and Barry Blake (eds.), *Handbook of Australian Languages*, Volume 2: 274–393. Amsterdam: John Benjamins.
- Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.
- Bresnan, Joan and Jonni M. Kanerva. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry* 20(1): 1–50.
- Bresnan, Joan and Sam A. Mchombo. 1986. Grammatical and anaphoric agreement. In *Papers from the Parasession on Pragmatics and Grammatical Theory*, Volume 22, Chicago, 278–292. Chicago Linguistic Society.
- Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63(4): 741–782.
- Bresnan, Joan and Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21(2): pp. 147–185.
- Broadwell, George A. 2006. *A Choctaw reference grammar*. Lincoln, NE: University of Nebraska Press.
- Bromley, Henry M. 1981. *A Grammar of Lower Grand Valley Dani*. Canberra: Australian National University.
- Brown, Robert. 1981. Semantic aspects of some Waris predications. In Franklin, Karl J. Ukarumpa (ed.), *Syntax and semantics in Papua New Guinea languages*: 93–123. Ukarumpa: Summer Institute of Linguistics.
- Brown, Robert. 1988. Waris case system and verb classification. *Language and Linguistics in Melanesia* 19: 37–80.
- Bubenik, Vit. 2006. Armenian. In Hewson, John and Vit Bubenik (eds.), *From Case to Adposition*: 160–177. Amsterdam-Philadelphia: John Benjamins.
- Buell, Leston C. 2005. *Issues in Zulu verbal morphosyntax*. Ph. D. thesis, UCLA.
- Butt, Miriam. 1993. Object specificity and agreement in Hindi/Urdu. In Beats K., G. Cooke, D. Kathman S. Kita K. McCullough and D. Testen (eds.), *Papers from the 29th Regional Meeting of the CLS*, Chicago, 89–104. Chicago Linguistic Society.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Cablitz, Gabriel H. 2006. *Marquesan: a grammar of space*. Trends in linguistics: Studies and monographs. Berlin-New York: Mouton de Gruyter.

- Campbell, Robert. 1977. Marcadores de fonte de informação na língua jama-madí. *Série Lingüística* 7(7): 117–125.
- Cardoso, Hugo. 2009. *The Indo-Portuguese language of Diu*. Utrecht.
- Carleton, Troi and Rachelle Waksler. 2000. Pronominal markers in Zenzontepec Chatino. *International Journal of American Linguistics* 66(3): 383–397.
- Carleton, Troi and Rachelle Waksler. 2002. Marking focus in Chatino. *Word* 53(2): 157–171.
- Caughley, Ross C. 1982. *The syntax and morphology of the verb in Chepang*. Pacific Linguistics, Series B. Canberra: Australian National University.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, Charles and Sandra Thompson (eds.), *Subject and Topic*: 25–56. London-New York: Academic Press.
- Chafe, Wallace. 1987. Cognitive constraints on information flow. In Tomlin, Russell (ed.), *Coherence and Grounding in Discourse*: 21–52. Amsterdam-Philadelphia: John Benjamins.
- Chamoreau, Claudine. 1999. Le marquage différentiel de l'objet en purépecha. *La Linguistique* 35(2): 99–114.
- Chamoreau, Claudine. 2000. *Grammaire du purépecha*. Lincom Studies in Native American Linguistics. Munich: Lincom Europa.
- Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chapman, Shirley. 1979. *Prominence in Paumari*. Archivo Linguístico. Brasília: Summer Institute of Linguistics.
- Chapman, Shirley and Desmond C. Derbyshire. 1991. Paumari. In Derbyshire, Desmond C. and Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages* 3: 161–352. Berlin-New York: Mouton de Gruyter.
- Chappell, Hilary. 2000. Dialect grammar in two early modern Southern Min texts: a comparative study of dative *kît*, comitative *câng*, and diminutive *-guia*. *Journal of Chinese Linguistics* 28(2): 247–302.
- Chappell, Hilary. 2006. From Eurocentrism to Sinocentrism: the case of disposal constructions in Sinitic languages. In Felix Ameka, Alan Dench and Nicholas Evans (eds.), *Catching language: the standing challenge of grammar writing*: 441–486. Berlin, New York: Mouton de Gruyter.
- Chappell, Hilary. to appear. Pan-Sinitic object markers: morphology and syntax. In Cao Guangshun, Hilary Chappell, R. Djamouri and Thekla Wiebusch (eds.), *Breaking down the barriers: interdisciplinary studies in Chinese linguistics and beyond*. Taipei: Academia Sinica.
- Chelliah, Shobhana. 1997. *A Grammar of Meithei*. Berlin-New York: Mouton de Gruyter.
- Chelliah, Shobhana. 2009. Semantic role to new information in Meithei. In Johanna, Barðdal and Shobhana Chelliah (eds.), *The role of semantics, pragmatics,*



- and discourse in the development of case: 377–400. Amsterdam-Philadelphia: John Benjamins.
- Chiriacescu, Sofiana and Klaus von Heusinger. 2010. Discourse prominence and *pe*-marking in Romanian. *International Review of Pragmatics* 2(2): 298–332.
- Chung, Sandra. 1978. *Case marking and grammatical relations in Polynesian*. Austin: University of Texas Press.
- Coelho, Gail. 2003. *A Grammar of Betta Kurumba*. Ph. D. thesis, University of Texas at Austin.
- Coghill, Eleanor. 2010. Ditransitive constructions in the Neo-Aramaic dialect of Telkepe. In Malchukov, Andrej, Martin Haspelmath and Bernard Comrie (eds.), *Studies in ditransitive constructions: a comparative handbook*: 221–242. Berlin-New York: Mouton de Gruyter.
- Coghill, Eleanor. to appear. Differential object marking in Neo-Aramaic. *Linguistics*. Special issue on *Differential object marking: theoretical and empirical issues*, Iemmolo Giorgio and Gerson Klumpp (eds.).
- Cohen, Marcel. 1936. *Traité de langue amharique (Abyssinie)*. Paris: Institut d'Ethnologie.
- Company Company, Concepción. 2002. Grammaticalization and category weakness. In Wischer, Ilse and Gabriele Diewald (eds.), *New reflections on grammaticalization*: 201–215. Amsterdam-Philadelphia: John Benjamins.
- Comrie, Bernard. 1973. The ergative: variations on a theme. *Lingua* 32: 239–253.
- Comrie, Bernard. 1975. The anti-ergative: Finland's answer to Basque. Volume 11, 112–121.
- Comrie, Bernard. 1979. Definite and animate objects: a natural class. *Linguistica Silesiana* 3: 15–21.
- Comrie, Bernard. 1989. *Language typology and linguistic universals. Syntax and morphology*. Oxford: Blackwell.
- Comrie, Bernard. 2003. When agreement gets trigger-happy. *Transactions of the Philological Society* 101: 313–337.
- Comrie, Bernard. 2008. Alignment of case marking of full noun phrases. In: Haspelmath, Martin, Matthew Dryer, David Gil, and Bernard Comrie (eds.). *The World Atlas of Language Structures Online*, chapter 98.
- Cook, Dorothy M. and Linda L. Criswell. 1993. *El idioma koreguaje (tucano occidental)*. Santafé de Bogotá, Columbia: Instituto Lingüístico de Verano.
- Cook, Dorothy M. and Stephen H. Levinsohn. 1985. Coreguaje: the domains of focus markers. In Brend, Ruth M. (ed.), 91–116. s 9. Dallas: Summer Institute of Linguistics. (ed.), *From phonology to discourse: studies in six Colombian languages*, Number 9 in Language Data, Amerindian Series: 91–116. Dallas/Arlington: Summer Institute of Linguistics.

- Cooreman, Ann. 1987. *Transitivity and discourse continuity in Chamorro narratives*. Berlin-New York: Mouton de Gruyter.
- Cooreman, Ann. 1988. Ergativity in Dyrbal discourse. *Linguistics* 26(5): 717–746.
- Coppock, Elizabeth and Stephen Wechsler. to appear. The objective conjugation in Hungarian: Agreement without phi features. *Natural Language and Linguistic Theory*.
- Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville. 2003. *Agreement*. Cambridge: Cambridge University Press.
- Creissels, Denis. 2005. A typology of subject and object markers in African languages. In Voeltz, Erhard F. K. (ed.), *Studies in African Linguistic Typology*: 43–70. Amsterdam-Philadelphia: John Benjamins.
- Creissels, Denis. ms. Transitive predication typology, alignment typology, word order typology. Talk given at the DGfS-CNRS Summer School on Linguistic Typology, Leipzig, August 15–September 3, 2010.
- Creissels, Denis and Jérémie Kouadio. 2010. Ditransitive constructions in Baule. In Malchukov, Andrej, Martin Haspelmath and Bernard Comrie (eds.), *Studies in ditransitive constructions: a comparative handbook*: 166–189. Berlin-New York: Mouton de Gruyter.
- Cristofaro, Sonia. 2005. *Subordination*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
- Cristofaro, Sonia. 2009. Grammatical categories and relations: universality vs. language-specificity and construction-specificity. *Language and Linguistics Compass* 3: 441–479.
- Cristofaro, Sonia. 2010. Semantic maps and mental representation. *Linguistic Discovery* 8(1): 35–52.
- Croft, William. 1988. Agreement vs. case marking in direct objects. In Barlow, Michael and Charles Ferguson (eds.), *Agreement in natural languages*: 159–179. Stanford: Center for the study of language and information.
- Croft, William. 1991. *Syntactic categories and grammatical relations: the cognitive organisation of information*. Chicago: University of Chicago Press.
- Croft, William. 2000. *Explaining language change: an evolutionary approach*. London: Longman.
- Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, William. 2010. Force dynamics and directed change in event lexicalization and argument realization. Ms. University of New Mexico, Albuquerque.
- Croft, William. to appear. *Verbs: aspect and argument structure*. Oxford: Oxford University Press.

- Cruschina, Silvio. 2006. Informational focus in Sicilian and the left periphery. In Frascarelli, Mara (ed.), *Phases of Interpretation*: 363–385. Berlin-New York: Mouton de Gruyter.
- Cruschina, Silvio. 2010. Syntactic extraposition and clitic resumption in Italian. *Lingua* 120(1): 50–73.
- Culy, Christopher. 2000. An incorporated topic marker in Takelma. In Butt, Miriam and Tracy Holloway King (eds.), *Proceedings of the LFG00 Conference*, Stanford. CSLI Publications.
- Curnow, Timothy J. 1997. *A Grammar of Awa Pit (Cuaiquer): An Indigenous Language of South-western Colombia*. Ph. D. thesis, Australian National University.
- Cyffer, Norbert. 1983. Case marking in Kanuri? *Africa und Übersee* 66: 191–202.
- Cysouw, Michael. 2003. Against implicational universals. *Linguistic Typology* 7(1): 89–101.
- Dabir Moghaddam, Mohammed. 1992. On the (in)dependence of syntax and pragmatics: evidence from the postposition *-râ* in Persian. In Stein, Dieter (ed.), *Cooperating with written texts: the pragmatics and comprehension of written texts*: 549–573. Berlin-New York: Mouton de Gruyter.
- Dahl, Östen. 2000a. Animacy and the notion of semantic gender. In Unterbeck, Barbara Matti Rissanen, Terttu Nevalainen and Mirja Saari (eds.), *Gender in grammar and cognition: approaches to gender*: 99–115. Berlin-New York: Mouton de Gruyter.
- Dahl, Östen. 2000b. Egophoricity in discourse and syntax. *Functions of Language* 7(1): 33–77.
- Dahl, Östen. 2008. Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua* 118(2): 141–150. Special issue on *Animacy, Argument Structure, and Argument Encoding*.
- Dahl, Östen and Kari Fraurud. 1996. Animacy in grammar and discourse. In Fretheim, Thorstein and Jeanette K. Gundel (eds.), *Reference and Referent Accessibility*: 47–64. Amsterdam-Philadelphia: John Benjamins.
- dal Jóhanna, Barð. 2001. *Case in Icelandic. A synchronic, diachronic and comparative approach*. Ph. D. thesis, Department of Scandinavian Languages, Lund University.
- Dalrymple, Mary and Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.
- Danon, Gabi. 2002. *Case and formal definiteness: the licensing of definite and indefinite noun phrases in Hebrew*. Ph. D. thesis, University of Tel Aviv.
- de Hoop, Helen and Andrej Malchukov. 2008. Case-marking strategies. *Linguistic Inquiry*: 39: 565–587.

- de Hoop, Helen and Bhuvana Narasimhan. 2008. Ergative case-marking in Hindi. In de Hoop, Helen and Peter de Swart (eds.), *Differential Subject Marking*: 63–78. Dordrecht: Kluwer.
- De Swart, Peter. 2006. Case markedness. In Kulikov Leonid, Andrej Malchukov and Peter de Swart (eds.), *Case, Valency and Transitivity*: 249–268. Amsterdam-Philadelphia: John Benjamins.
- De Swart, Peter. 2007. *Cross-linguistic Variation in Object Marking*. Utrecht: LOT Publishers.
- Dede, Keith. 2007. The origin of anti-ergative *xa* in Huangshui Chinese. *Language and Linguistics* 8(4): 863–881.
- DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 57(3): 626–657.
- DeLancey, Scott. 1984. Notes on agentivity and causation. *Studies in Language* 8(1): 181–213.
- Demuth, Katherine and Mark Johnson. 1989. Interactions between discourse functions and agreement in Setawana. *Journal of African Languages and Linguistics* 11(1): 22–35.
- Demuth, Katherine A. 1989. Discourse functions of independent pronouns in Setswana. In Haïk, Isabelle and Laurice Truller (eds.), *Current approaches to African linguistics*: 176–193. Dordrecht: Foris.
- Den Besten, Hans. 2000. The slaves' language in the Dutch Cape Colony and Afrikaans *vir*. *Linguistics* 38: 949–971.
- Derbyshire, Desmond C. 1986. Comparative survey of morphology and syntax in Brazilian Arawakan. In Derbyshire, Desmond C. and Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages 1*: 469–566. Berlin-New York: Mouton de Gruyter.
- Diez, Friedrich. 1863. *Introduction à la grammaire des langues romanes*. Paris: Librairie A. Franck.
- Dimmendaal, Gerrit. 2009. Tama. In Dimmendaal, Gerrit (ed.), *Coding participant marking*: 307–331. Amsterdam-Philadelphia: John Benjamins.
- Dimmendaal, Gerrit. 2010. Differential object marking in Nilo-Saharan. *Journal of African Languages and Linguistics* 13(1): 13–46.
- Dixon, Robert M. W. 1979. Ergativity. *Language*: 59–138.
- Dixon, Robert M. W. 1981. Wargamay. In Dixon, Robert M. W. and Barry Blake (eds.), *Handbook of Australian languages*, Volume 2: 1–144. Canberra: Australian National University.
- Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Dixon, Robert M. W. and Alan R. Vogel. 2004. *The Jarawara language of Southern Amazonia*. Oxford linguistics. Oxford: Oxford University Press.
- Donohue, Mark. 2002. Tobati. In Lynch, John, Malcolm Ross and Terry Crowley (eds.), *The Oceanic Languages*: 186–203. London: Routledge.

- Donohue, Mark and Søren Wichmann. 2008. *The typology of semantic alignment*. Oxford: Oxford University Press.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3): 547–619.
- Dryer, Matthew. 1986. Primary objects, secondary objects, and antitativity. *Language* 62(4): 808–845.
- Dryer, Matthew. 1989. Large linguistic areas and language sampling. *Studies in Language* 13: 257–292.
- Dryer, Matthew. 1992. The Greenbergian word order correlations. *Language* 68: 81–138.
- Dryer, Matthew. 1996. Grammatical relations in Ktunaxa (Kutenai): The Belcourt Lecture delivered before the University of Manitoba on 24 february 1995.
- Dryer, Matthew. 1997. Are grammatical relations universal? In Joan Bybee, John Haiman and Sandra Thompson (eds.), *Essays on language function and language type: dedicated to Talmy Givón*: 115–143. Amsterdam: John Benjamins.
- Dryer, Matthew. 2008. Order of subject, object and verb. In Haspelmath, Martin, Dryer Matthew, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures Online*: Munich: Max Planck Digital Library.
- Du Bois, John. 1987. The discourse basis of ergativity. *Language* 63(4): 805–855.
- Du Bois, John. 2003. Argument structure: grammar in use. In Dubois, John, Lorraine E. Kumpf and William J. Ashby (eds.), *Preferred argument structure*: 11–60. Amsterdam-Philadelphia: John Benjamins.
- Dum-Tragut, Jasmine. 2009. *Armenian: modern Eastern Armenian*. London Oriental and African language library. Amsterdam-Philadelphia: John Benjamins.
- Dumitrescu, Domnita. 1997. El parámetro discursivo en la expresión del objeto directo lexical: español madrileño vs. español porteño. *Signo y Señal* 7: 305–354.
- Dumitrescu, Domnita. 1998, 140-152. "a" personal, duplicación clítica y marcidez: español porteño vs. español madrileño. In M, Ward A. (ed.), *Actas del XII Congreso de la Asociación Internacional de Hispanistas*, Birmingham. University of Birmingham.
- Duranti, Alessandro. 1979. Object clitic pronouns in Bantu and the Topicality Hierarchy. *Studies in African Linguistics* 10(1): 31–45.
- Duranti, Alessandro and Elinor Ochs. 1979. Left-dislocation in Italian conversation. In Givón, Talmy (ed.), *Syntax and semantics 12: Discourse and syntax*: 377–416. New York: Academic Press.
- Elwolde, John. 1994. The use of 'et in non-Biblical Hebrew texts. *Vetus Testamentum* 1(1): 170–182.
- Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1): 1–25.

- Engsheden, Åke. 2008. Differential object marking in Sahidic Coptic. In Josephson, Folke and Ingmar Söhrman (eds.), *Interdependence of Diachronic and Synchronic Analyses*: 323–344. Amsterdam-Philadelphia: John Benjamins.
- Epps, Patience. 2008. *A Grammar of Hup*. Mouton Grammar Library. Berlin-New York: Mouton de Gruyter.
- Epps, Patience. 2009. Where differential object marking and split plurality intersect: evidence from Hup. In Epps, Patience and Alexandre Arkhipov (eds.), *New Challenges in Typology*: 85–104. Berlin-New York: Mouton de Gruyter.
- Erguvanli, Eser E. 1984. *The function of word order in Turkish grammar*. Berkeley: University of California Press.
- Ernout, Alfred and Antoine Meillet. 1985. *Dictionnaire étymologique de la langue latine: histoire des mots* (4 ed.). Paris: Klincksieck.
- Escandell-Vidal, Victoria. 2007. Topics from Ibiza: differential object marking and clitic-dislocation. In Kaiser, Georg and Manuel Leonetti (eds.), *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*, Konstanz, 23–43. Fachbereich Sprachwissenschaft der Universität Konstanz.
- Escandell-Vidal, Victoria. 2009. Differential object marking and topicality: the case of Balearic Catalan. *Studies in Language* 33(4): 832–884.
- Ezard, Bryan. 1997. *A Grammar of Tawala, an Austronesian language of the Milne Bay Area, Papua New Guinea*. Pacific Linguistics, Series C. Canberra: Australian National University.
- Farina, Margherita. 2006. La marca differenziale dell'oggetto e la particella 'et in ebraico biblico, prime prospezioni. In *Quaderni del Laboratorio di Linguistica*, Volume 5:
- Farrell, Patrick. 2005. *Grammatical Relations*. Oxford: Oxford University Press.
- Farrell, Tim. 1995. Fading ergativity? A study of ergativity in Balochi. In Bennet, David C., Theodora Bynon and George Hewitt (eds.), *Subject, voice and ergativity: Selected essays*: 218–243. London: School of African and Oriental Studies.
- Feldman, Harry. 1986. *A grammar of Awtuw*. Canberra: Australian National University.
- Fiedler, Ines. 2009. Contrastive topic marking in Gbe. In *Current Issues in Unity and Diversity of Languages. Collection of papers selected from the CIL 18*, Seoul, 295–308. The Linguistic Society of Korea.
- Filimonova, Elena. 2005. The noun phrase hierarchy and relational marking: problems and counterevidence. *Linguistic Typology* 9(1): 77–113.
- Fillmore, Charles. 1968. The case for case. In Bach, Emmon and Robert Harms (eds.), *Universals in Linguistic Theory*: 1–88. New York: Holt, Rinehart and Winston.

- Finer, Daniel L. 1997. Contrasting A'-dependencies in Selayarese. *Natural Language and Linguistic Theory* 15(4): 677–728.
- Foley, William. 1986. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley, William. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29(1): 357–404.
- Foley, William. 2007. Typology of information packaging in the clause. In Shopen, Timothy (ed.), *Language Typology and Syntactic description*, Volume 3: 362–446. Cambridge: Cambridge University Press.
- Foreman, Velma. 1974. *Grammar of Yessan-Mayo*. Language Data, Asian-Pacific Series. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Formentin, Vittorio. 2004. Quattro note sintattiche dal "Tristano Veneto". In Dardano, Maurizio and Gianluca Frenguelli (eds.), *SintAnt. La sintassi dell'italiano antico. Atti del Convegno Internazionale di studi*: 175–196. Roma: Aracne.
- Fried, Robert W. 2011. *A grammar of Bao'an Tu, a Mongolic language of Northwest China*. Ph. D. thesis, State University of New York at Buffalo.
- Gabriel, Christoph and Esther Rinke. 2010. Information packaging and the rise of clitic-doubling in the history of Spanish. In Ferraresi, Gisella and Rosemarie Lühr (eds.), *Diachronic Studies on Information Structure. Language Acquisition and Change*: 63–86. Berlin-New York: Mouton de Gruyter.
- García García, Marco. 2007. Differential object marking with inanimate objects. In Leonetti, Manuel and Georg Kaiser (eds.), *Proceedings of the workshop "Definiteness, specificity and animacy in Ibero-Romance languages"*, Konstanz, 63–84. Fachbereich Sprachwissenschaft, Universität Konstanz.
- Genetti, Carol. 1994. *A Descriptive and Historical Account of the Dolakha Newari Dialect (Monumenta Serindica)*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Genetti, Carol. 1997. Object relations and dative case in Dolakha Newari. *Studies in Language* 21(1): 33–62.
- Genetti, Carol. 2007. *A Grammar of Dolakha Newari*. Berlin-New York: Mouton de Gruyter.
- Gerner, Matthias. 2008. Ambiguity-driven differential object marking in Yon-gren Lolo. *Lingua* 118(3): 296–331.
- Ghomeshi, Jila. 1997. Topics in Persian VPs. *Lingua* 102(1): 133–167.
- Givón, Talmy. 1976. Topic, pronoun and grammatical agreement. In Li, Charles (ed.), *Subject and topic*: 149–188. London-New York: Academic Press.
- Givón, Talmy. 1978. Definiteness and referentiality. In Greenberg, Joseph (ed.), *Universals of human language*, Volume 4: 291–330. Stanford: Stanford University Press.
- Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

- Givón, Talmy. 1983. Topic continuity in discourse: an introduction. In Givón, Talmy (ed.), *Topic continuity in discourse. A quantitative cross-language study*: 1–43. Amsterdam-Philadelphia: John Benjamins.
- Givón, Talmy. 1985. Iconicity, isomorphism and non-arbitrary coding in syntax. In Haiman, John (ed.), *Iconicity in syntax*: 187–219. Amsterdam-Philadelphia: John Benjamins.
- Givón, Talmy. 1992. The grammar of referential coherence as mental processing instructions. *Linguistics* 30(1): 5–56.
- Givón, Talmy. 1997. Grammatical relations: an introduction. In Givón, Talmy (ed.), *Grammatical Relations: a functionalist perspective*: 1–84. Amsterdam-Philadelphia: John Benjamins.
- Givón, Talmy. 2001. *Syntax. A functional-typological introduction*, Volume 2. Amsterdam-Philadelphia: John Benjamins.
- Glinert, Lewis. 1989. *The grammar of Modern Hebrew*. New York: Cambridge University Press.
- Gorelova, Liliya M. 2002. *Manchu grammar*. Handbook of oriental studies. Leiden: Brill.
- Goudswaard, Nelleke E. 2005. *The Begak (Ida'an) Language of Sabah*. Utrecht: LOT Publishers.
- Greenberg, Joseph. 1966. *Language Universals, with special reference to feature hierarchies*. Janua Linguarum, Series Minor 59. The Hague: Mouton.
- Gregores, Emma and Jorge A. Suárez. 1967. *A Description of Colloquial Guaraní*. The Hague: Mouton.
- Grunow-Hårsta, Karen. 2004. Direction and differential dative case marking in Magar. In Saxena, Anju (ed.), *Himalayan Languages*: 77–100. Berlin-New York: Mouton de Gruyter.
- Güldemann, Tom. 2007. Preverbal objects and information structure in Benue-Congo. In Aboh, Enoch O., Katharina Hartmann and Malte Zimmermann (eds.), *Focus strategies in African languages: the interaction of focus and grammar in Niger-Congo and Afro-Asiatic*, Trends in Linguistics. Studies and Monographs: 83–111. Berlin-New York: Mouton de Gruyter.
- Güldemann, Tom. forthcoming-a. Morphology: !Xõo of Lone Tree. In Voßen, Rainer (ed.), *The Khoisan languages*, Routledge Language Family Series: London: Routledge.
- Güldemann, Tom. forthcoming-b. Morphology: !Xam of Strandberg. In Voßen, Rainer (ed.), *The Khoisan languages*, Routledge Language Family Series: London: Routledge.
- Güldemann, Tom. forthcoming-c. Syntax: !Xam of Strandberg. In Voßen, Rainer (ed.), *The Khoisan languages*, Routledge Language Family Series: London: Routledge.



- Güldemann, Tom. forthcoming-d. Syntax: !Xõo of Lone Tree. In Voßen, Rainer (ed.), *The Khoisan languages*, Routledge Language Family Series: London: Routledge.
- Gulya, János. 1966. *Eastern Ostyak Chrestomathy*. Bloomington: Indiana University Press.
- Gunlogson, Christine. 2001. Third-person object prefixes in Babine-Witsuwit'en. *International Journal of American Linguistics* 67(4): 365–395.
- Guntsetseg, Dolgor. 2008. Differential object marking in Mongolian. In Schäfer, Florian (ed.), *SinSpeC Working papers of the SFB 732 "Incremental specification in Context"*: Stuttgart. University of Stuttgart.
- Haiman, John. 1978. Conditionals are topics. *Language* 54(3): 564–589.
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4): 781–819.
- Haiman, John and Paola Benincà. 1992. *The Rhaeto-Romance languages*. Romance Linguistics. London: Routledge.
- Harris, Alice C.. 2002. *Endoclitics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.
- Harrison, David. 2000. *Topics in the phonology and morphology of Tuvan*. Ph. D. thesis, Yale University.
- Haspelmath, Martin. 20008. Ditransitive constructions: the verb "give". In Haspelmath, Martin, Matthew Dryer David Gil and Bernard Comrie (eds.), *The World Atlas of Language Structures Online*: Munich: Max Planck Digital Library.
- Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11(1): 119–132.
- Haspelmath, Martin. 2008. Parametric versus functional explanations of syntactic universals. In Biberauer, Theresa (ed.), *The limits of syntactic variation*: 75–107. Amsterdam-Philadelphia: John Benjamins.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 3(3): 663–687.
- Havers, Wilhelm. 1925. Der sogenannte "Nominativus Pendens". *Indogermanische Forschungen* 43: 207–256.
- Hawkins, John. 1978. *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. London: Croom Helm.
- Hawkinson, Annie and Larry Hyman. 1974. Hierarchies of natural topic in Shona. *Studies in African Linguistics* 5(1): 147–170.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph. D. thesis, University of Massachusetts, Amherst.
- Heine, Bernd and Tania Kouteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

- Hellenthal, Anneke C. 2010. *A grammar of Sheko*. Ph. D. thesis, Leiden University.
- Hercus, Louise A. 1994. *A Grammar of the Arabana-Wangkangurru Language, Lake Eyre Basin, South Australia*. Pacific Linguistics, Series C. Canberra: Australian National University.
- Heusinger, Klaus von and Georg Kaiser. 2011. Affectedness and differential object marking in Spanish. *Morphology* 21(1).
- Hildebrandt, Kristine A. 2004. A grammar and dictionary of the Manange language. In Genetti, Carol (ed.), *Tibeto-Burman languages of Nepal: Manange and Sherpa*: 2–189. Canberra: Pacific Linguistics.
- Hills, E. C.. 1920. The accusative "a". *Hispania* 3(4): 216–222.
- Himmelman, Nikolaus. 2005. The Austronesian languages of Asia and Madagascar: typological characteristics. In Adelaar, Alexander and Nikolaus Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*: 110–173. London: Routledge.
- Hollenbach, Barbara. 1992. A syntactic sketch of Copala Trique. In Bradley, Henry C. and Barbara E. Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*: 173–431. Dallas: Summer Institute of Linguistics.
- Holmer, Nils Magnus. 1983. *Linguistic survey of South-Eastern Queensland*. Pacific Linguistics D-54. Canberra: Australian National University.
- Hopper, Paul and Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56: 251–299.
- Hopper, Paul and Elizabeth Traugott. 1993. *Grammaticalization* (1st ed.). Cambridge: Cambridge University Press.
- Hume, Elizabeth. 2004. Deconstructing markedness: a predictability-based account. In *Berkeley Linguistics Society*, BLS, Berkeley.
- Hutchison, John P. 1980. The Kanuri associative postposition: a case for subordination. *Studies in African linguistics* 11(3): 321–351.
- Hutchison, John P. 1981. *The Kanuri language: a reference grammar*. Madison: African Studies Program, University of Wisconsin.
- Hutchison, John P. 1986. Major constituent case marking in Kanuri. In Dimmendaal, Gerrit (ed.), *Current approaches to African linguistics*: 191–208. Dordrecht: Foris.
- Huumo, Tuomas. 2003. Incremental existence: The world according to the Finnish existential sentence. *Linguistics* 41(3): 461–493.
- Huumo, Tuomas. 2009. Fictive dynamicity, nominal aspect, and the Finnish copulative construction. *Cognitive Linguistics* 20(1): 43–70.
- Hyman Larry, M. and Alessandro Duranti. 1982. On the object relation in Bantu. In Hopper, Paul and Sandra Thompson (eds.), *Syntax and semantics: studies in transitivity*, Volume 15: 217–239. New York: Academic Press.

- Iemmolo, Giorgio. 2009. La marcatura differenziale dell'oggetto in siciliano antico. *Archivio Glottologico Italiano* 94(2): 185–225.
- Iemmolo, Giorgio. 2010a. La marcatura differenziale dell'oggetto in siciliano: un'analisi contrastiva. In Iliescu, M. et al. (ed.), *Actes du XXV Congrès International de Linguistique et de Philologie Romanes*, Volume 2, Berlin-New York, 341–350. Mouton de Gruyter.
- Iemmolo, Giorgio. 2010b. Topicality and differential object marking: evidence from Romance and beyond. *Studies in Language* 34(2): 239–272.
- Iemmolo, Giorgio. 2011. Towards a typological study of differential object marking and differential object indexation. Talk given at the "9th Association for Linguistic Typology Biennial Meeting", Hong Kong, July 2011.
- Iemmolo, Giorgio. in preparation. Differential vs. split object marking: a functional differentiation.
- Imart, Guy. 1981. *Le Kirghiz. Description d'une langue de littérisation récente*. Aix-en-Provence: Publications de l'Université de Provence.
- Ingram, Andrew. 2001. *A grammar of Anamuxra: a Language of Madang province, Papua New Guinea*. Ph. D. thesis, University of Sydney.
- Ismatullaev, Khairulla and Walter Feldman. 1995. *Modern literary Uzbek*. Bloomington: Indiana University, Research Institute for Inner Asian Studies.
- Jacobs, Joachim. 2001. The dimensions of topic-comment. *Linguistics* 39: 641–681.
- Jake, Janice L. 1980. Object verb agreement in Tigre. *Studies in the linguistic sciences* 10(1): 71–84.
- Jakobi, Angelika. 1989. *A Fur grammar*. Hamburg: Buske Verlag.
- Jakobi, Angelika. 2009. Differential object marking in Uncuwee (Kordofan Nubian). In *Paper presented at the 6th World Congress of African Languages, Cologne, August 17-19, 2009*.
- Janse, Mark. 2008. Clitic doubling from Ancient to Asia Minor Greek. In Kallulli, Dalina and Liliane Tasmowski (eds.), *Clitic Doubling in the Balkan Languages*: 165–202. Amsterdam-Philadelphia: John Benjamins.
- Jansen, Joanna. 2010. *A grammar of Yakima Ichishkiin/Sahaptin*. Ph. D. thesis, University of Oregon.
- Jauncey, Dorothy. 2002. Tamabo. In Lynch, John, Malcolm Ross, and Terry Crowley (eds.), *The Oceanic Languages*: 608–625. London: Routledge.
- Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language and Linguistic Theory* 2(1): 39–76.
- Jenny, Matthias. 2009. Differential object marking in Burmese. Talk given at the workshop "Differential object marking: theoretical and empirical issues", Helsinki, August 2009.
- Jing-Schmidt, Zhuo. 2005. *Dramatized discourse: the Mandarin Chinese ba-construction*. Amsterdam-Philadelphia: John Benjamins.

- Joly, André. 1971. Le complément verbal et le morphème "a" en béarnais. observation sur le genre et la fonction dans les langues romanes. *Zeitschrift für romanische Philologie* 87: 286–305.
- Jones, Wendell and Paula Jones. 1991. *Barasano Syntax*, Volume 101 of *Publications in Linguistics*. Dallas: Summer Institute of Linguistics and The University of Texas at Arlington.
- Josephs, Lewis S. 1975. *Palauan Reference Grammar*. Honolulu: University of Hawaii Press.
- Kachru, Yamuna. 2006. *Hindi*. Amsterdam-Philadelphia: John Benjamins.
- Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions: *râ* in Persian. *Linguistic Analysis* 3-4: 139–191.
- Karlsson, Fred. 2003. *Finnish: an essential grammar*. London: Routledge.
- Karttunen, Lauri. 1976. Discourse referents. In McCawley, James (ed.), *Syntax and Semantics 7: notes from the linguistic underground*: 363–385. New York: Academic Press.
- Keenan, Edward. 1976. Some universals of passives in universal grammar. In *Papers from the XIth Annual Meeting of the Chicago Linguistic Society*, Chicago, 340–352. Chicago Linguistic Society.
- Keenan, Edward. 2008. The definiteness of subjects and objects in Malagasy. In Corbett, Greville and Michael Noonan (eds.), *Case and grammatical relations: studies in honor of Bernard Comrie*, Volume 81 of *Typological Studies in Language*: 241–261. Amsterdam-Philadelphia: John Benjamins.
- Keenan, Edward L. and Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1): pp. 63–99.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam-Philadelphia: John Benjamins.
- Khan, Geoffrey. 1984. Object markers and agreement pronouns in Semitic languages. *Bulletin of the School of Oriental and African Studies* 47(3): 468–500.
- Khan, Geoffrey. 2008. *The Neo-Aramaic dialect of Barwar. Grammar*. Handbuch der Orientalistik: Nahe und der Mittlere Osten. Brill.
- Kievit, Dirk and Saliem Kievit. 2009. Differential object marking in Tigrinya. *Journal of African Languages and Linguistics* 30(1): 45–71.
- Kifle, Nazareth Amlesom. 2007. Differential object marking and topicality in Tigrinya. In King, Tracy Holloway and Miriam Butt (eds.), *The Proceedings of the LFG07 Conference*, Stanford, 5–25. CSLI Publications.
- Kimball, Geoffrey. 1991. *Koasati Grammar*. Lincoln, NE: University of Nebraska Press.
- Kiss, Katalin É. 2011. Null pronominal objects in Hungarian: a case of exap-tation. In *Presentation given at the workshop "The diachrony of referential null arguments" at the International Conference on Historical Linguistics, Osaka, July 2011*.

- Kittilä, Seppo. 2002. *Transitivity. Towards a comprehensive typology*. Ph. D. thesis, University of Turku, Finland.
- Kittilä, Seppo and Andrej Malchukov. 2008. Varieties of accusative. In Malchukov, Andrej and Andrew Spencer (eds.), *Handbook of Case*: 549–561. Oxford: Oxford University Press.
- Klamer, Marian. 2008. Differential marking of intransitive subjects in Kambara (Austronesian). In de Hoop, Helen and Peter de Swart (eds.), *Differential Subject Marking*: 281–299. Dordrecht: Kluwer.
- Klamer, Marian. 2010. *A grammar of Teiwa*. Berlin-New York: Mouton de Gruyter.
- Klamer, Marian and František Kratochvíl. 2006. The role of animacy in Teiwa and Abui. In *Berkeley Linguistics Society*, Volume 32, Berkeley. Berkeley Linguistic Society.
- Klumpp, Gerson. 2009. Variation in Komi object marking. In Dufter, Andreas, Jürg Fleischer and Guido Seiler (eds.), *Describing and modeling variation in grammar*, Trends in Linguistics, Studies and Monographs 204: 325–359. Berlin-New York: Mouton de Gruyter.
- König, Christa. 2009. *Case in Africa*. Oxford: Oxford University Press.
- König, Ekkehard. 2001. Intensifiers and reflexive pronouns. In Haspelmath, Martin, Ekkehard König Wulf Oesterreicher and Wolfgang Raible (eds.), *Language universals and language typology: an international handbook*: 747–760. Berlin-New York: Mouton de Gruyter.
- Korn, Agnes. 2008. Marking of arguments in Balochi. ergative and mixed constructions. In Karimi, Simin Vida Samiian and Donald Stilo (eds.), *Aspects of Iranian linguistics*: 249–276. Newcastle: Cambridge Scholars Publishing.
- Korn, Agnes and Carina Jahani. 2009. Balochi. In Windfuhr, Gernot (ed.), *The Iranian languages*: 634–692. London: Routledge.
- Kornfilt, Jaklin. 1997. *Turkish*. Descriptive Grammars. London: Routledge.
- Körtvély, Erika. 2005. *Verb conjugation in Tundra Nenets*. Studia Uralo-Altaica. Szeged: SZTE Finnugor Tanszék.
- Koutsoudas, Andreas. 1967. Object particles in Lebanese. *Journal of the American Oriental Society* 87(4): 512–517.
- Kratochvíl, František. 2007. *A Grammar of Abui*. Ph. D. thesis, Leiden University, Leiden.
- Kratochvíl, František. under revision. Differential case marking in Abui. *Linguistics*.
- Krifka, Manfred. 2007. Basic notions of information structure. In Fery, C. and Manfred Krifka (eds.), *Interdisciplinary studies of information structure*, Volume 6: University of Potsdam.
- Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge grammatical descriptions. Cambridge: Cambridge University Press.

- Kuno, Susumu. 1987. *Functional syntax: anaphora, discourse and empathy*. Chicago: The University of Chicago Press.
- Kuno, Susumu and Etsuki Kubaraki. 1977. Empathy and syntax. *Linguistic Inquiry* 8(4): 627–672.
- La Polla, Randy L. 1992. Anti-ergative marking in Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 15(1): 1–9.
- La Polla, Randy L. 1993. On the change to verb-medial word order in Proto-Chinese: evidence from Tibeto-Burman. In H. Kitamura, T. Nishida and Y. Nagano (eds.), *Current Issues in Sino-Tibetan Linguistics*: 98–104. Osaka: National Museum of Ethnology.
- La Polla, Randy L. 1994. Parallel grammaticalizations in Tibeto-Burman: evidence of sapir's drift. *Linguistics of the Tibeto-Burman Area* 17(1): 61–80.
- La Polla, Randy L. 1995. Pragmatic relations and word order in Chinese. In Downing, Pamela A. and Michael Noonan (eds.), *Word order in discourse*: 297–329. Amsterdam-Philadelphia: John Benjamins.
- La Polla, Randy L. 2004. On nominal relational morphology in Tibeto-Burman. In Ying-jin Lin, Fang-min Hsu, Chun-chih Lee Jackson T.-S. Sun Hsiu-fang Yang and Dah an Ho (eds.), *Papers in Honor of Professor Hwang-Cherng Gong on his Seventieth Birthday*: 43–74. Taipei: Institute of Linguistics, Academia Sinica.
- La Polla, Randy L. and Dory Poa. 2006. On describing word order. In Ameka, Felix, Alan Dench and 2006. Nicholas Evans, 269–295. Berlin: Mouton de Gruyter (eds.), *Catching language: the standing challenge of grammar writing*, Volume 167 of *Trends in linguistics. Studies and monographs*: 269–295. Berlin-New York: Mouton de Gruyter.
- Laca, Brenda. 1995. Sobre el uso del acusativo preposicional en español. In Pensado, Carmen (ed.), *El complemento directo preposicional*: 61–91. Madrid: Visor Libros.
- Laca, Brenda. 2006. El objeto directo. La marcación preposicional. In Company-Company, Concepción (ed.), *Sintaxis histórica del español. Primera parte: La frase verbal*, Volume 1 of *Lengua y Estudios Literarios*: 423–475. Mexico: Fondo de cultura económica y Universidad Nacional Autónoma de México.
- Lahaussois, Aimée. 2002. *Aspects of the grammar of Thulung Rai: an endangered Himalayan language*. Ph. D. thesis, University of California at Berkeley, Berkeley.
- Lakoff, George. 1977. Linguistic gestalts. In *Chicago Linguistic Society*, Number 13, Chicago, 236–287. Chicago University Press.
- Lambrecht, Knud. 1994. *Information structure and sentence form: topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press.

- Lambrecht, Knud. 2001. Dislocation. In Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), *Language universals and language typology: an international handbook*: 1050–1078. Berlin-New York: Mouton de Gruyter.
- Langacker, Ronald. 1991. *Foundations of cognitive grammar, descriptive application*, Volume 2. Stanford: Stanford University Press.
- Layton, Bentley. 2000. *A Coptic grammar with chrestomathy and glossary: Sahidic dialect*. Porta Linguarum Orientalium. Wiesbaden: Harrassowitz.
- Lazard, Gilbert. 1994. *L'actance*. Paris: Presses Universitaire de France.
- Lazard, Gilbert. 2001. Le marquage différentiel de l'objet. In Haspelmath, Martin, Ekkehard König Wulf Oesterreicher and Wolfgang Raible (eds.), *Language universals and language typology: an international handbook*: 873–855. Berlin-New York: Mouton de Gruyter.
- Lazard, Gilbert. 2003. What is an object in a crosslinguistic perspective? In Fiorentino, Giuliana (ed.), *Romance Objects. Transitivity in Romance Languages*: 1–16. Berlin-New York: Mouton de Gruyter.
- Lazard, Gilbert. 2006. *La quête des invariants interlangues: La linguistique est-elle une science?* Paris: Champion.
- Lehman, Thomas. 1993. *A grammar of Modern Tamil*. Pondicherry: Institute of Language and Culture.
- Lehmann, Christian. 1982. Universal and typological aspects of agreement. In Seiler, Hansjacob and Franz Josef Stachowiak (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen: II: Die Techniken und ihr Zusammenhang in Einzelsprachen*: 201–267. Tübingen: Narr.
- Lehmann, Christian. 1988. On the function of agreement. In Barlow, Michael and Charles Ferguson (eds.), *Agreement in Natural Language*: 55–67. Stanford: CSLI Publications.
- Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: Lincom Europa.
- Leonetti, Manuel. 2004. Specificity and differential object marking in Spanish. *Catalan Journal of Linguistics* 3(1): 75–114.
- Leonetti, Manuel. 2008. Specificity in clitic doubling and in differential object marking. *Probus* 20(1): 35–69.
- Leslau, Wolf. 1995. *Reference grammar of Amharic*. Wiesbaden: Harrassowitz.
- Levin, Aryeh. 1987. The particle *la* as a object marker in some Arabic dialects of the Galilee. *Journal of Arabic Linguistics* 17(1): 31–40.
- Levin, Beth and Malka Rappaport Hovav. 2005. *Argument realization*. Oxford: Oxford University Press.
- Leyew, Zelealem. 2003. *The Kemantney language. A sociolinguistic and grammatical study of language replacement*. Cologne: Rüdiger Köppe Verlag.

- Li, Charles and Sandra Thompson (eds.). 1976. *Subject and topic*. London-New York: Academic Press.
- Li, Charles N. and Sandra Thompson. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
- Li, Yen-Hui Audrey. 1990. *Order and constituency in Mandarin Chinese*. Dordrecht: Kluwer.
- Li, Yen-Hui Audrey. 2006. Chinese *ba*. In Martin, Everaert and Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, Volume 1: 374–468. Oxford: Blackwell.
- Lidz, Jeffrey. 2006. The grammar of accusative case in Kannada. *Language* 82(1): 1–23.
- Liu, Feng-Hsi. 1997. An aspectual analysis of *ba*. *Journal of East Asian Linguistics* 6(1): 51–99.
- Liu, Feng-Hsi. 2007. Word order variation and *ba* sentences in Chinese. *Studies in Language* 31(3): 649–682.
- López Cruz, José Gerardo. 2008. Las construcciones bitransitivas en el trique de San Juan Copala. Master's thesis, Universidad de Sonora, Mexico.
- Lord, Carol. 1982. The development of object markers in serial verb languages. In Hopper, Paul and Sandra Thompson (eds.), *Syntax and semantics*, Volume 15: 277–299. Academic Press.
- Lord, Carol. 1993. *Historical change in serial verb constructions*. Amsterdam-Philadelphia: John Benjamins.
- Loughnane, Robyn. 2009. *A Grammar of Oksapmin*. Ph. D. thesis, University of Melbourne.
- Louwerse, John. 1988. *The morphosyntax of Una in relation to discourse structure*. Pacific Linguistics, Series B. Canberra: Australian National University.
- Lukas, Johannes. 1953. *Die Sprache der Tubu in der zentralen Sahara*, Volume 14 of *Deutsche Akademie der Wissenschaften Institut Orientforschung*. Berlin: Akademie Verlag.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- MacDonald, Lorna. 1990. *A grammar of Tauya*. Berlin-New York: Mouton de Gruyter.
- Mahootian, Shahrzad. 1997. *Persian*. London: Routledge.
- Malchukov, Andrej. 2005. Case pattern splits, verb types and construction competition. In Amberber, Mengistu and Helen de Hoop (eds.), *Competition and variation in natural languages: the case for case*: 73–118. Amsterdam: Elsevier.
- Malchukov, Andrej. 2006. Transitivity parameters and transitivity alternations: constraining co-variation. In Kulikov Leonid, Andrej Malchukov and Peter de Swart (eds.), *Case, valency and transitivity*: 329–359. Amsterdam-



- Philadelphia: John Benjamins.
- Malchukov, Andrej. 2008. Animacy and asymmetries in differential case marking. *Lingua* 118: 203–221.
- Marcantonio, Angela. 1981. Topic-focus and some typological consideration in Hungarian. *Acta Linguistica Scientiarum Hungaricae* 31: 219–235.
- Marcantonio, Angela. 1985. On the definite vs. indefinite conjugation in Hungarian: a typological and diachronic analysis. *Acta Linguistica Scientiarum Hungaricae* 35: 267–298.
- Marie, Nilsenova.. 2002. The pragmatics of differential object marking in Western Armenian. In Nissim, Malvina (ed.), *Proceedings of the ESSLLI '02 Student Session*.
- Marten, Lutz, Nancy C. Kula and Nhlanhla Thwala. 2007. Parameters of morphosyntactic variation in Bantu. *Transactions of the Philological Society* 105(1): 1–86.
- Martins, Silvana and Valteir Martins. 1999. The Maku language family. In Dixon, Robert M. W. and Alexandra Y. Aikhenvald (eds.), *Amazonian languages*: 251–267. Cambridge: Cambridge University Press.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4: 307–333.
- Maslova, Elena. 2003a. A case for implicational universals. *Linguistic Typology* 7(1): 101–108.
- Maslova, Elena. 2003b. *A Grammar of Kolyma Yukaghir*. Berlin-New York: Mouton de Gruyter.
- Maslova, Elena. 2003c. Information focus in relational clause structure. In Tsunoda, Tasaku and Taro Kageyama (eds.), *Voice and grammatical relations. Festschrift for Masayoshi Shibatani*: 175–194. Amsterdam-Philadelphia: John Benjamins.
- Matisoff, James A. 1973. *A grammar of Lahu*. University of California Publications in Linguistics. Berkeley: University of California Press.
- Matras, Yaron. 2002. *Romani: a linguistic introduction*. Cambridge: Cambridge University Press.
- McCloskey, James and Kenneth Hale. 1983. On the syntax of person-number inflection in Modern Irish. *Natural Language and Linguistic Theory* 1(4): pp. 487–533.
- McGregor, R.S. 1972. *Outline of Hindi grammar*. Oxford: Oxford University Press.
- Mchombo, Sam A. 2004. *The syntax of Chichewa*. Cambridge Syntax Guides. Cambridge: Cambridge University Press.

- McLendon, Sally. 1975. *A Grammar of Eastern Pomo*, Volume 74. Berkeley: University of California Press.
- McLendon, Sally. 1978. Ergativity, case, and transitivity in Eastern Pomo. *International Journal of American Linguistics* 1(1): 1–9.
- Megerdoomian, Karine. 1999. Projection of direct objects. New Mexico State University.
- Meier, Harri. 1947. O problema do acusativo preposicional no Catalão. *Boletim de Filologia* 8: 237–260.
- Meillet, Antoine. 1903. *Esquisse d'une grammaire comparée de l'arménien classique*. Impr. des PP. Mékhitharistes.
- Melis, Chantal. 1995. El objeto directo personal en El Cantar de Mio Cid. estudio sintáctico-pragmático. In Pensado, Carmen (ed.), *El complemento directo preposicional*: 133–163. Madrid: Visor Libros.
- Meyer-Lübke, Wilhelm. 1900. *Grammaire des langues romanes. Syntaxe*, Volume 3. Paris: H. Welter.
- Miller, Marion. 1999. *Desano Grammar*. Studies in the languages of Colombia. Dallas/Arlington: Summer Institute of Linguistics.
- Mira Mateus, Maria H., Ana M. Brito Inês Duarte and Faria Hub. 1989. *Gramática da língua portuguesa*. Lisboa: Caminho.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4): 847–894.
- Mithun, Marianne. 1991. Active/agentive case marking and its motivations. *Language* 67(3): 510–546.
- Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.
- Mithun, Marianne. 2008. The emergence of agentive systems. In Donohue, Mark and Søren Wichmann (eds.), *The typology of semantic alignment systems*: 297–333. Oxford: Oxford University Press.
- Mithun, Marianne and Wallace Chafe. 1999. What are S, A and O? *Studies in Language* 23: 569–596.
- Mohanan, Tara. 1994. *Argument structure in Hindi*. Stanford: CSLI Publications.
- Molnár, Valeria. 2001. Contrast from a contrastive perspective. In Kruijff-Korbayová, Irina and M. Steedman (eds.), *Information structure, discourse structure and discourse semantics, ESSLLI2001 Workshop Proceedings*, Helsinki, 99–114. The University of Helsinki Press.
- Moravcsik, Edith. 1974. Object-verb agreement. *Working Papers on Language Universals* 15: 25–40.
- Moravcsik, Edith. 1978. On the case marking of objects. In Greenberg Joseph, Charles Ferguson and Edith Moravcsik (eds.), *Universals of human language. Syntax*, Volume 4: 249–290. Stanford: Stanford University Press.

- Morimoto, Yukiko. 2002. Prominence mismatches and differential object marking in Bantu. In Butt, Miriam and Tracy Holloway King (eds.), *Proceedings of the LFG-02 Conference*, Stanford. CSLI Publications.
- Morse, Nancey L. and Michael B. Maxwell. 1999. *Cubeo grammar*. Studies in the Languages of Colombia. Arlington: University of Texas and Summer Institute of Linguistics.
- Moyse-Faurie, Claire. 2004. Convergence entre thème et focus dans les langues polynésiennes. *Faits de Langues* 23-24: 303–320.
- Murane, Elizabeth. 1974. *Daga grammar: from morpheme to discourse*. Norman: Summer Institute of Linguistics of the University of Oklahoma.
- Næss, Åshild. 2004. What markedness marks: the markedness problem with direct objects. *Lingua* 114: 1186–1212.
- Næss, Åshild. 2007. *Prototypical transitivity*. Amsterdam-Philadelphia: John Benjamins.
- Næss, Åshild. 2008. Varieties of dative. In Malchukov, Andrej and Andrew Spencer (eds.), *Handbook of Case*: 572–580. Oxford: Oxford University Press.
- Neuburger, Kathrin Ann and Elisabeth Stark. submitted. Differential object marking in Corsican: regularities, triggering factors, functions.
- Neukom, Lukas. 2000. Argument marking in Santali. *Mon-Khmer Studies* 30(1): 95–113.
- Newmeyer, Frederick J. 2005. *Possible and probable languages: a generative perspective on linguistic typology*. Oxford: Oxford University Press.
- Nichols, Lynn. 1997. *Topics in Zuni syntax*. Ph. D. thesis, Harvard University.
- Niculescu, Alexandru. 1959. Sur l'objet direct prépositionnel dans les langues romanes. In *Recueil d'études romanes, publié à l'occasion du IXème Congrès International de linguistique romane à Lisbonne*: 167–185. Bucarest.
- Nikolaeva, Irina. 1999. Object agreement, grammatical relations, and information structure. *Studies in Language* 23: 331–376.
- Nikolaeva, Irina. 2001. Secondary topic as a relation in information structure. *Linguistics* 39: 1–49.
- Nikolaeva, Irina and Maria Tolskaya. 2001. *A grammar of Udihe*. Mouton Grammar Library. Berlin: Mouton de Gruyter.
- Nocentini, Alberto. 1985. Sulla genesi dell'oggetto preposizionale nelle lingue romanze. In Nocentini, Alberto (ed.), *Studi filologici e linguistici per Carlo Alberto Mastrelli*: 299–311. Pisa: Pacini.
- Noonan, Michael. 1991. Anti-dative shift. *Milwaukee Studies on Language* 5: 50–58.
- Noonan, Michael. 1992. *A grammar of Lango*. Mouton Grammar Library. Berlin-New York: Mouton de Gruyter.
- Nuger, Justin. 2007. A minimalist analysis of Palauan differential object marking and object agreement. Ms. UC Santa Cruz.

- Nuger, Justin. 2009. On downward-entailing existentials and differential object marking in Palauan. In Chung, Sandra, Daniel Finer Ileana Paul and Eric Potsdam (eds.), *Proceedings of AFLA 16*, London, Ont, 137–151. AFLA Proceedings Project.
- Ochs Keenan, Elinor and Bambi Schieffelin. 1976a. Foregrounding referents: a reconsideration of left dislocation in discourse. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, Volume 2, 240–257.
- Ochs Keenan, Eleanor and Bambi Schieffelin. 1976b. Topic as a discourse notion: a study of topic in the conversations of children and adults. In Li, Charles (ed.), *Subject and Topic*: 337–385. New York: Academic Press.
- Olson, Michael L. 1981. *Barai clause junctures: toward a functional theory of interclausal relations*. Ph. D. thesis, Australian National University, Canberra.
- Osumi, Midori. 1995. *Tinrin Grammar*. Honolulu: University of Hawaii Press.
- Pandharipande, Rajeshwari. 1997. *Marathi*. Descriptive grammars. London: Routledge.
- Parry, Mair. 2003. L'oggetto preposizionale nel ligure medievale. *Verbum. Analecta Neolatina* 5(1): 113–126.
- Paul, Waltraud. 2002. Sentence-internal topics in Mandarin Chinese: the case of object preposing. *Language and Linguistics* 3(4): 695–714.
- Pearson, Joel. 2001. *The clause structure of Malagasy: a minimalist approach*. Ph. D. thesis, UCLA.
- Pensado, Carmen. 1995. La creación del objeto directo preposicional y la flexión de los pronombres personales en las lenguas románicas. In Pensado, Carmen (ed.), *El complemento directo preposicional*: 179–233. Madrid: Visor Libros.
- Perlmutter, David. 1983. *Studies in Relational Grammar I*, Volume 1. Chicago: University of Chicago Press.
- Peterson, David A. 2007. *Applicative constructions*. Oxford studies in typology and linguistic theory. Oxford: Oxford University Press.
- Peyraube, Alain. 1985. Les structures en *ba* en chinois medieval et moderne. *Cahiers de Linguistique Asie-Orientale* (1): 193–213.
- Peyraube, Alain. 1994. Nouvelles réflexions sur l'histoire des formes accusatives en *ba* du chinois. *Cahiers de Linguistique Asie-Orientale* 23(1): 265–277.
- Peyraube, Alain. 1996. Recent issues in Chinese historical syntax. In Huang, C.-T. James and Yen-Hui Audrey Li (eds.), *New horizons in Chinese linguistics*: 161–214. Dordrecht: Kluwer.
- Pilot-Raichoor, Christiane. 1994. L'objet en badaga. *Bulletin de la Société de Linguistique de Paris* 89(1): 359–397.
- Plank, Frans (ed.). 1984. *Objects. Towards a theory of grammatical relations*. London and New York: Academic Press.
- Plungian, Vladimir. 1995. *Dogon*. Languages of the world materials. Munich: Lincom Europa.

- Poppe, Nicholas. 1960. *Buriat Grammar*. Bloomington: Indiana University.
- Post, Mark. 2007. *A grammar of Galo*. Ph. D. thesis, La Trobe University Research Centre for Linguistic Typology.
- Primus, Beatrice. 1999. *Cases and thematic roles. Ergative, accusative and active*. Tübingen: Narr.
- Primus, Beatrice. to appear. Animacy, generalized semantic roles, and differential object marking. In Lamers, Monique and Peter de Swart (eds.), *Case, word order, and prominence: psycholinguistic and theoretical approaches to argument structure*: Dordrecht: Springer Netherlands.
- Prince, Ellen. 1981. Toward a taxonomy of given/new information. In Cole, Peter (ed.), *Radical Pragmatics*: 223–254. Academic Press.
- Prince, Ellen. 1997. On the functions of left-dislocation in English discourse. In Kamio, A (ed.), *Directions in Functional Linguistics*: 117–144. Amsterdam-Philadelphia: John Benjamins.
- Putzu, Ignazio E. 2008. Per uno studio dell'accusativo preposizionale in sardo antico: emergenze dallo spoglio del *Condaghe di San Pietro di Silki*. In Lazzeroni, Romano, Emanuele Banfi Giuliano Bernini Marina Chini and Giovanna Marotta (eds.), *Diachronica et synchronica: studi in onore di Anna Giacalone Ramat*: 397–428. Pisa: ETS.
- Ramirez, Henri. 1997. *A fala tukano dos ye'pâ-masa. Gramática*. Manaus: Inspeitoria Salesiana Missionária da Amazônia, CEDEM.
- Rasch, Jeffrey W. 2002. *The basic morpho-syntax of Yaitepec Chatino*. Ph. D. thesis, Rice University.
- Rice, Keren. 1989. *A grammar of Slave*. Mouton Grammar Library. Berlin-New York: Mouton de Gruyter.
- Rice, Keren. 2003. Doubling by agreement in Slave (Northern Athapaskan). In Carnie, Andrew, Heidi Harley and Mary Ann Willie (eds.), *Formal approaches to function in grammar. In honor of Eloise Jelinek*, Linguistik Aktuell/Linguistics Today: 51–78. Amsterdam-Philadelphia: John Benjamins.
- Riedel, Kristina. 2009. *The syntax of object marking in Smbaa: a comparative Bantu perspective*. Ph. D. thesis, Leiden University.
- Rijkhoff, Jan, Dik Bakker Kees Hengeveld and Peter Kahrel. 1993. A method of language sampling. *Studies in Language* 17(1): 169–203.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Haegeman, Liliane (ed.), *Elements of grammar*: Dordrecht: Kluwer.
- Rohlf, Gerhard. 1969. *Grammatica storica della lingua italiana e dei suoi dialetti. Sintassi e formazione delle parole*, Volume 3. Torino: Einaudi.
- Rohlf, Gerhard. 1971/1984. Autour de l'accusatif prépositionnel dans les langues romanes. (Concordances et Discordances). In Rohlf, Gerhard (ed.), *Von Rom zur Romània*: 61–80. Tübingen: Narr.

- Rosch, Eleanor. 1978. Principles of categorization. In Eleanor, Rosch and B. B. Lloyd (eds.), *Cognition and categorization*: 27–48. Hillsdale, NJ: Erlbaum.
- Rounds, Carol. 2009. *Hungarian: an essential grammar*. Routledge Essential Grammars. Taylor and Francis.
- Rude, Noel. 1982. Promotion and topicality of Nez Perce objects. *Berkely Linguistic Society* 8: 463–483.
- Rude, Noel. 1986. Topicality, transitivity, and the direct object in Nez Perce. *International Journal of American Linguistics* 52: 124–153.
- Rude, Noel. 1988. Semantic and pragmatic objects in Klamath. In Shipley, William (ed.), *In honour of Mary Haas*: 651–673. Berlin-New York: Mouton de Gruyter.
- Rude, Noel. 1994. Direct, inverse and passive in Northwest Sahaptin. In Givón, Talmy (ed.), *Voice and Inversion*, Volume 28 of *Typological Studies in Language*: 101–119. Amsterdam-Philadelphia: John Benjamins.
- Rude, Noel. 1997a. Dative shifting and double objects in Sahaptin. In Givón, Talmy (ed.), *Grammatical Relations*, Volume 35 of *Typological Studies in Language*: Typological Studies in Language. John Benjamins.
- Rude, Noel. 1997b. On the history of nominal case in Sahaptian. *International Journal of American Linguistics* 63(1): 113–143.
- Russi, Cinzia. 2008. *Italian clitics: an empirical study*. Berlin-New York: Mouton de Gruyter.
- Sapir, Edward. 1909. *Takelma Texts*. University of Pennsylvania Anthropological Publications.
- Sapir, Edward. 1922. The Takelma language of Southwestern Oregon. In Boas, Franz (ed.), *Bulletin of american ethnology: Handbook of American Indian Languages*, Volume 40: 1–296. Washington, DC: Washington, DC: Government Printing Office.
- Satti, Nasir Ali Osman. 2008. *A grammatical analysis of Dongolese phrases and clauses*. Ph. D. thesis, University of Khartoum.
- Savà, Graziano. 2005. *A grammar of Ts'amakko*. Cushitic Language Studies. Cologne: Rüdiger Köppe Verlag.
- Saxon, Leslie. 1989. Lexical versus syntactic projection: the configurationality of Slave. In Cook, Eung-Doand Keren Rice (ed.), *Athapaskan Linguistics: current perspectives on a language family*, Volume 15 of *Trends in Linguistics*: 379–406. Berlin-New York: Mouton de Gruyter.
- Schiffman, Harold. 1999. *A reference grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
- Schneider-Blum, Gertrud. 2007. *A grammar of Alaaba, a Highland East Cushitic language of Ethiopia*. Köln: Rüdiger Köppe Verlag.
- Seidl, Amanda and Alexis Dimitriadis. 1997. The discourse function of object marking in Swahili. In Singer, Kora, Randall Eggert and Gregory Anderson

- (eds.), *Chicago Linguistic Society 33: papers from the Main Session*, Chicago. Chicago Linguistic Society.
- Seiler, Walter. 1985. *Imonda, a Papuan Language*. Pacific Linguistics, Series B. Canberra: Australian National University.
- Shain, Cory and Judith Tonhauser. 2010. The synchrony and diachrony of differential object marking in Paraguayan Guaraní. *Language Variation and Change* 22(3): 321–346.
- Siewierska, Anna. 1984. *The passive: a comparative linguistic analysis*. London: Croom Helm.
- Siewierska, Anna. 1988. *Word order rules*. London: Croom Helm.
- Siewierska, Anna. 1997. The formal realization of case and agreement marking: a functional perspective. In Simon-Vandenberg, Anne-Marie, Kristin Davidse and Dirk Noël (eds.), *Reconnecting Language*. Amsterdam: John Benjamins, 181–210. Amsterdam-Philadelphia: John Benjamins.
- Siewierska, Anna. 1999. From anaphoric pronoun to grammatical agreement marker: why objects don't make it. *Folia Linguistica* 33(1-2): 225–251.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Siewierska, Anna. 2010. From 3ptl-to passive: incipient, emergent and established passives. *Diachronica* 27(3): 73–109.
- Siewierska, Anna and Dik Bakker. 2009. Case and alternative strategies. In Malchukov, Andrej and Andrew Spencer (eds.), *Handbook of Case*: 290–303. Oxford: Oxford University Press.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Dixon, Robert M. W (ed.), *Grammatical categories in Australian languages*: 112–171. Canberra: Australian National University.
- Sinnemäki, Kaius. submitted. Differential and consistent case marking of object: a typological study. *Linguistics*.
- Skrubnik, Elena. 2001a. Pragmatic structuring in Northern Mansi. In Seilenthal, Tõnu (ed.), *Congressus Nonus Internationalis Fenno-ugristarum. Dissertationes sectionum: Linguistica III*, Volume 6, Tartu, 222–239. University of Tartu.
- Skrubnik, Elena. 2001b. Variation of noun phrase markers in Siberian languages. In Boeder, Winfried and Gerd Hentsche (eds.), *Variierende Markierung von Nominalgruppe in Sprachen unterschiedlichen Typs*: 345–364. Oldenburg: Oldenburg, Bibliotheks- und Informationssystem der Universität Oldenburg.
- Slater, Keith W. 2003. *A grammar of Mangghuer: a Mongolic language of China's Qinghai-Gansu Sprachbund*. London: Routledge.
- Smeets, Ineke. 2008. *A grammar of Mapuche*. Mouton grammar library. Berlin-New York: Mouton de Gruyter.
- Song, Jae Jung. 2001. *Linguistic typology: morphology and syntax*. Harlow and London: Pearson Education.

- Sornicola, Rosanna. 2002. Concezioni del parlato e concezioni del mutamento linguistico. Ricordando Monica Berretta. *Italian Journal of Linguistics* 14(2): 375–390.
- Soukka, Maria. 2000. *A Descriptive Grammar of Noon*, Volume 40 of *Lincom Studies in African Linguistics*. Munich: Lincom Europa.
- Spitz, Walter. 2002. Voice and role in two Philippine languages. In Wouk, Fay and Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, *Pacific Linguistics* 518: 379–404. Canberra: Pacific Linguistics.
- Stenzel, Kristine. 2008. Kotiria "differential object marking" in cross-linguistic perspective. *Amerindia* 32(3/4): 153–181.
- Stilo, Donald. 2010. Ditransitive constructions in Vafsi: a corpus-based study. In Malchukov, Andrej, Martin Haspelmath and Bernard Comrie (eds.), *Studies in ditransitive constructions: a comparative handbook*: 243–276. Berlin-New York: Mouton de Gruyter.
- Strom, Clay. 1992. *Retuarã Syntax*. Studies in the languages of Colombia. Dallas: Summer Institute of Linguistics.
- Stucky, Susan. 1981. *Word order variation in Makua: a phrase structure grammar analysis*. Ph. D. thesis, University of Illinois at Urbana-Champaign.
- Stump, Gregory and Ramawater Yadav. 1988. Maithili verb agreement and the control agreement principle. In Diane Brentari, Gary N. Larson and Lynn A. MacLeod (eds.), *CLS 24: Papers from the 24th Annual Regional Meeting of the Chicago Linguistic Society: Part II: Parasession on Agreement in Grammatical Theory*, Chicago, 304–321. Chicago Linguistic Society.
- Sun, Chao-Fen. 1996. *Word order change and grammaticalization in the history of Chinese*. Stanford: Stanford University Press.
- Sun, Chao-Fen and Talmy Givón. 1985. On the so-called SOV word order in Mandarin Chinese: a quantified text study and its implications. *Language* 61(2): 329–351.
- Talmy, Leonard. 1976. Semantic causative types. In Shibatani, Masayoshi (ed.), *The grammar of causative constructions*: 43–116. London-New York: Academic Press.
- Talmy, Leonard. 1988. Force dynamics in language and cognition. *Cognitive Science* 12:49-100. 12(1): 49–100.
- Tamm, Anne. 2005. *Relations between Estonian verbs, aspect, and case*. Ph. D. thesis, Eötvös Loránd University, Budapest.
- Teng, S.F.C. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Pacific linguistics, Series C. Canberra: Australian National University.
- Terrill, Angela. 2002. *Dharumbal: the language of Rockhampton, Australia*, Volume 525 of *Pacific Linguistics*. Canberra: Australian National University.



- Thomas, Earl W.. 1969. *The syntax of spoken Brazilian Portuguese*. Nashville: Vanderbilt University Press.
- Thompson, Alexander. 1909. Beiträge zur Kasuslehre I. *Indogermanische Forschungen* 24: 293–307.
- Thompson, Chad. 1990. On the treatment of topical objects in Chepang: passive or inverse? *Studies in Language* 14(2): 405–427.
- Thomson, Robert W.. 1989. *An introduction to classical Armenian*. Caravan Books.
- Thurgood, Graham. 1978. Thematicization and aspects of the verbal morphology in Burmese: the principles of organization. *PBLS* (4): 254–67.
- Timberlake, Alan. 1977. Reanalysis and actualization in syntactic change. In Li, Charles (ed.), *Mechanisms of syntactic change*: 141–180. Austin: University of Texas Press.
- Tomić, Olga. 2006. *Balkan Sprachbund morpho-syntactic features*. Dordrecht: Springer.
- Torrego Salcedo, Ester. 1999. El complemento directo preposicional. In Bosque, Ignacio and Violeta De Monte (eds.), *Gramàtica descriptiva de la Lengua Española*, Volume 2: 1779–1805. Madrid: Espasa.
- Tosco, Mauro. 1994. On case marking in the Ethiopian language area (with special reference to subject marking in East Cushitic). In Brugnatelli, Vermondo (ed.), *Sem Cam Iafet. Atti della 7a Giornata di Studi Camito-Semitici e Indeu-ropei*, Milano, 225–244. Centro Studi Camito-Semitici.
- Tosco, Mauro. 2010. Why contrast matters: information structure in Gawwada (East Cushitic). In Fiedler, Ines and Anne Schwarz (eds.), *The expression of information structure: a documentation of its diversity across Africa*: 315–347. Amsterdam-Philadelphia: John Benjamins.
- Traugott, Elizabeth. 1995. Subjectification in grammaticalization. In Stein, Dieter and Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*: 31–54. Cambridge: Cambridge University Press.
- Traugott, Elizabeth and Richard Dasher. 2005. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Tsao, Feng-Fu. 1987. A topic-comment approach to the *ba*-construction. *Journal of Chinese Linguistics* 1: 1–54.
- Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21: 385–396.
- Tupper, Ian. 2009. Transitivity and noun phrase marking in Namia. *La Trobe Papers in Linguistics* 12.
- Underriner, Janne. 2002. *Intonation and syntax in Klamath*. Ph. D. thesis, University of Oregon.
- Vallduvì, Enric. 1992. *The informational component*. New York: Garland.

- van der Voort, Hein. 2004. *A grammar of Kwaza*. Mouton Grammar Library. Berlin-New York: Mouton de Gruyter.
- van der Wal, Jenneke. 2009. *Word order and information structure in Makhuwa-Enahara*. Ph. D. thesis, Leiden University.
- van Valin, Robert and Randy La Polla. 1997. *Syntax*. Cambridge: Cambridge University Press.
- Virtanen, Susanna. submitted. Pragmatic direct object marking in Eastern Mansi.
- Vogel, Alan R. 1989. Gender and gender agreement in Jaruára (Arauan). Master's thesis, University of Texas at Arlington.
- von Heusinger, Klaus. 1997. *Salienz und Referenz. Der Epsilonoperator in der Semantik der Nominalphrase und anaphorischer Pronomen*. Berlin: Akademie Verlag.
- von Heusinger, Klaus. 2002. The cross-linguistic implementations of specificity. In Jaszczolt, Katarzyna and Ken Turner (eds.), *Meanings in Contrast: The Cambridge Papers*: 405–421. Amsterdam-Philadelphia: John Benjamins.
- von Heusinger, Klaus. to appear. Specificity. In Heusinger, Klaus von, Claudia Maienborn and Paul Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*, Volume 2: Berlin-New York: Mouton de Gruyter.
- von Heusinger, Klaus and Georg Kaiser. 2005. The evolution of differentiated object marking in Spanish. In Stark, Elisabeth, Klaus von Heusinger and Georg Kaiser (eds.), *Proceedings of the workshop "Specificity and the Evolution/Emergence of Nominal Determination Systems in Romance"*, Konstanz, 33–69. Fachbereich Sprachwissenschaft, Universität Konstanz.
- von Heusinger, Klaus and Georg Kaiser. 2007. Differential object marking and the lexical semantics of verbs in Spanish. In Leonetti, Manuel and Georg Kaiser (eds.), *Proceedings of the workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*, Konstanz, 83–111. Fachbereich Sprachwissenschaft, Universität Konstanz.
- von Heusinger, Klaus and Jaklin Kornfilt. 2005. The case of the direct object in Turkish: semantics, syntax, morphology. *Turkic Languages* 9(1): 3–44.
- Waag, Christine. 2009. *The Fur verb and its context*. Ph. D. thesis, University of Cologne.
- Wald, Benji. 1979. The development of the Swahili object marker: a study of the interaction of syntax and discourse. In Givón, Talmy (ed.), *Syntax and semantics: discourse and syntax*, Volume 12: 505–524. New York: Academic Press.
- Wald, Benji. 1997. Varbrul and the human/inanimate polarization of the Swahili object marker. In Guy Gregory R., Crawford Feagin, Deborah Schiffrin and John Baugh (eds.), *Towards a Social Science of Language*: 311–338. Amsterdam-Philadelphia: John Benjamins.
- Wali, K. and O.N. Koul. 1997. *Kashmiri: a cognitive-descriptive grammar*. Descriptive grammars. London: Routledge.

- Waltke, Bruce and Michael P. O'Connor. 1990. *An introduction to Biblical Hebrew syntax*. Winona Lake: Eisenbrauns.
- Wang, L.. 1947. *Modern Chinese Grammar*. Shanghai: Zhonghua Shuju.
- Wantalangi, Max. 1993. *The Manadonese grammar*. Ph. D. thesis, La Trobe University.
- Watters, David E. 2002. *A grammar of Kham*. Cambridge: Cambridge University Press.
- Watters, John R. 2003. Grassfields Bantu. In Nurse, Derek and Gérard Philippson (eds.), *The Bantu languages: 225–256*. London: Routledge.
- Weiss, Doris. 2009. *Phonologie et morphosyntaxe du maba*. Ph. D. thesis, Université Lumière Lyon 2.
- Weissenrieder, Maureen. 1991. A functional approach to the accusative A. *Hispania* 74(1): 146–156.
- Wheeler, Alva. 1970. *Grammar of the Siona Language, Colombia, South America*. Ph. D. thesis, University of California at Berkeley.
- Windfuhr, Gernot. 1979. *Persian grammar: history and state of its study*. Trends in Linguistics. State-of-the-Art Reports. Mouton de Gruyter.
- Windfuhr, Gernot. 1990. Persian. In Comrie, Bernard (ed.), *The world's major languages: 523–546*. London: Croom Helm.
- Windfuhr, Gernot and John R. Perry. 2009. Persian and Tajik. In Windfuhr, Gernot (ed.), *The Iranian languages: 416–544*. London: Routledge.
- Woldemariam, Hirut. 2004. *A grammar of Haro with comparative notes on the Ometo linguistic group*. Ph. D. thesis, Addis Ababa University.
- Woldemariam, Hirut. 2009. Haro. In Dimmendaal, Gerrit (ed.), *Coding participant marking: 97–122*. Amsterdam-Philadelphia: John Benjamins.
- Woolford, Ellen. 2000. Object agreement in Palauan: Specificity, humanness, economy and optimality. In Paul, Ileana, Vivianne Phillips and Lisa Travis (eds.), *Formal Issues in Austronesian Linguistics: 215–245*. Dordrecht: Kluwer.
- Wurm, Stephen. 1976. Accusative marking in Duungidjawa (Waga-Waga). In Dixon, R.M.W. (ed.), *Grammatical categories in Australian languages: 106–111*. Canberra: Australian Institute of Aboriginal Studies.
- Yadav, Ramawater. 1996. *A reference grammar of Maithili*. Trends in linguistics: Documentation. Berlin-New York: Mouton de Gruyter.
- Yamamoto, Mutsumi. 1999. *Animacy and reference: a cognitive approach to corpus linguistics*. Amsterdam-Philadelphia: John Benjamins.
- Yang, Ning. 2008. *The indefinite object in Mandarin Chinese: its marking, interpretation and acquisition*. Utrecht: LOYT Publishers.
- Yang, Ning and Geertie van Bergen. 2007. Scrambled objects and case marking in Mandarin Chinese. *Lingua* 9: 1617–1635.
- Yip, Po-Ching and Don Rimmington. 2004. *Chinese: a comprehensive grammar*. London: Routledge.

- Zribi-Hertz, Anne and Liliane Mbolatianavalona. 1999. Towards a modular theory of linguistic deficiency: evidence from Malagasy personal pronouns. *Natural Language and Linguistic Theory* 17(1): 161–218.
- Zúñiga, Fernando. 2006. *Deixis and alignment. Inverse systems in indigenous languages of the Americas*. Amsterdam-Philadelphia: John Benjamins.
- Zúñiga, Fernando. 2007. The discourse-syntax interface in Northwestern Amazonia. Differential object marking in Makú and some Tucanoan languages. In Wetzels, Leo (ed.), *Language endangerment and endangered languages: linguistic and anthropological studies with special emphasis on the languages and cultures of the Andean-Amazonian Border Area*: 209–227. Leiden University: Publications of the Research School of Asian, African, and Amerindian Studies.
- Zúñiga, Fernando. 2010. La marca diferencial del objeto en mapudungún. *Lingüística* 24: 141–164.