DE GRUYTER MOUTON Linguistic Typology 2025; 29(1): 127–165

Svenja Bonmann*, Sonja Riesberg and Nikolaus P. Himmelmann

Differential object marking in Western Malayo-Polynesian symmetrical voice languages

https://doi.org/10.1515/lingty-2024-0002
Received January 17, 2024; accepted July 19, 2024; published online September 11, 2024

Abstract: Differential object marking (DOM) is known from a great variety of language families. We present data suggesting that, contrary to what is suggested in recent typological surveys, DOM is also a widespread phenomenon in Austronesian symmetrical voice languages. To do this, we show why the identification of DOM constructions in these languages is difficult, propose a definition for DOM that takes note of the peculiarities of symmetrical voice systems, and briefly review three types of DOM systems attested in the area. Our findings have implications for claims regarding the universality of DOM, and quality control in compiling large-scale typological databases.

Keywords: differential object marking; Austronesian; symmetrical voice; referential scales; prominence

1 Introduction

Differential object marking (DOM) is commonly understood as the phenomenon where direct objects, roughly defined as the patient-like non-subject core argument of transitive verbs, allow for (at least) two kinds of morphosyntactic marking depending on factors such as animacy, definiteness, specificity, or information structure.[1] DOM is known from several hundred languages of unrelated language

1 Major references include Silverstein (1976), Comrie (1979), Moravcik (1978), Croft (1988), Bossong (1985), Aissen (2003), de Swart (2007), Dalrymple and Nikolaeva (2011), Witzlack-Makarevich and Seržant (2018), Kalin (2018), Bárány and Kalin (2020) and Kagan (2020) among others.

***Corresponding author: Svenja Bonmann [ˈsvɛn.ja ˈbɔn.man]**, University of Cologne, Cologne, Germany, E-mail: svenja.bonmann@uni-koeln.de. https://orcid.org/0000-0002-4005-7891
Sonja Riesberg [ˈzɔn.ja ˈʁiːs.b̥ɛɐ̯kʰ], LaCiTO – CNRS, Villejuif, France; and University of Cologne, Cologne, Germany, E-mail: sonja.riesberg@cnrs.fr. https://orcid.org/0000-0002-0252-0910
Nikolaus P. Himmelmann [ˈnɪ.ko.laʊ̯s ˈhɪ.məl.man], University of Cologne, Cologne, Germany, E-mail: sprachwisssenschaft@uni-koeln.de. https://orcid.org/0000-0002-4385-8395

families and branches of larger phyla and, for this reason, has been widely considered a robustly attested cross-linguistic phenomenon in need of a general explanation. Here, we call this the universalist view on DOM. This view has recently been challenged by Bickel et al. (2015) who argue that "differential case marking on A and P is first and foremost a pattern prone to diffusion" (2015: 40). We call this here the contact view on DOM (i.e. DOM emerges and is maintained in contact settings; see also Rodríguez-Ordóñez 2017 or Mardale and Karatsareas 2020). Bickel et al.'s argument has been carefully examined by Schmidtke-Bode and Levshina (2018) regarding methodological, conceptual and theoretical issues, and found wanting in all these regards. Using the same data set as Bickel et al., they find support for the universalist view, i.e. that "there *is* evidence for universal scale effects on case marking" (Schmidtke-Bode and Levshina 2018: 531).

The present contribution concerns another problematic issue, which is not specific to Bickel et al.'s study, but rather pertains to all work using large-scale cross-linguistic databases. Compiling such databases involves making decisions on whether to include a particular phenomenon and, if so, how to code it in accordance with the guidelines for the database. Various kinds of problems occur in this process, ranging from simple typos when keying in a data point, to systematic misrepresentations of the structures attested in a language or group of languages. The latter is the issue of interest in the present contribution, Western Malayo-Polynesian (WMP) languages being our primary example. These languages have been largely neglected in the DOM literature, and when they are mentioned at all, they are classified inconsistently, sometimes being considered DOM languages, sometimes not. Our goal here is to show that such inconsistencies occur, and why they occur.

We should emphasize from the outset that this article is not about mistakes (such as typos) that have been made in compiling the databases we use as our main examples here. As further illustrated in Sections 2 and 3, there are of course different ways of defining DOM constructions and of constraining the data to be included in a database for methodological or pragmatic reasons. Such differences are related to the specific research question and the resources available in a particular project. Our concern here is the fact that evaluating data from WMP languages when compiling a DOM database is complicated, among other things, by the fact that these languages have multiple basic transitive constructions (also known as symmetrical voice systems), which in turn renders difficult the task of deciding whether a given construction can be considered a DOM construction (see Section 4). In as much as our argument that DOM is robustly attested among WMP symmetrical voice languages is convincing, it provides support for the universalist view of DOM. It shows that DOM is attested in an area which so far has been widely considered to lack DOM constructions (see Section 6).

The structure of this article is as follows: Section 2 summarizes major DOM patterns, using Iranian DOM languages for illustration. The motivation for this is to make clear that all the phenomena we document for DOM in WMP languages are also found in prototypical DOM languages such as Iranian ones, which have repeatedly served as a major point of reference in the DOM literature (e.g., Bickel et al. 2015; Bossong 1985; Haig 2018; Iemmolo 2011: 162–167). Sections 3 and 4 introduce the general typological characteristics of WMP languages, and their implications for identifying DOM constructions. Section 5 provides examples for DOM in WMP symmetrical voice languages and classifies them into several subtypes according to the specificity scale. Section 6 concludes with a discussion of our findings.

2 Differential object marking – some basic distinctions as found in Iranian languages

In line with much of the literature, we consider DOM to pertain to direct objects which allow for (at least) two kinds of morphosyntactic marking depending on factors such as animacy, specificity or topicality (see Section 2.2). A direct object is defined as the more patient-like core argument of a prototypical two-place transitive predicate, the morphosyntactic marking of which differs from the marking of the single core argument of an intransitive predicate (cf. Sinnemäki 2014: 287). Furthermore, the morphosyntactic marking distinction should be specific to direct objects and not pertain to all core argument functions.

In Section 2.1, the following basic distinctions regarding DOM are briefly illustrated: (i) the distinction between DOM and differential object indexing (DOI); (ii) the fact that DOM may involve case marking or adpositions; (iii) the distinction between symmetric and asymmetric DOM; and (iv) the fact that diagnosing DOM involves comparing marking differences across different parts of speech, most importantly nouns and pronouns. Section 2.2 briefly introduces the referential scales relevant for determining when objects occur with differential marking.

2.1 Major morphosyntactic distinctions

Several strategies for the differential marking of direct objects have been described in the literature. They include case marking (e.g., Iron Ossetic, Yaghnobi, Turkish), marking by an adposition (e.g., Spanish, Hindi, and several Pamir languages including Sanglechi), clitic doubling or pronominal reduplication (e.g., Romanian, Catalan), and agreement (e.g., Swahili). When the differential marking occurs on the

verb, as in cases of clitic doubling, it is also common to speak of differential object indexing (DOI; Compensis 2022; Haig 2018). DOI occurs in some Austronesian languages, especially in the Oceanic branch and in southern Sulawesi (see Iemmolo 2011: 203–206 and Appendix 1), but we are only concerned with DOM proper in this article.

The two most common marking strategies are DOM by means of case marking and DOM via adpositions. For instance, in Parachi the preposition *ma* marks specific direct objects (Kieffer 2009: 699), non-specific direct objects remaining unmarked, see example (1). Otherwise, *ma* marks a) a specific location, time, and goal (e.g., *ma dur* 'at the river', *ma dōwās* 'at 12 o'clock', *ma [žū qāter] suwār nhōšt* 'he sat down on/mounted [a mule]'), or b) recipients/addressees (as in *[ma guɣōn] āo dhaym* 'I shall give water [to the cows]').

(1) Parachi, Iranian, DOM by means of an adposition (Morgenstierne 1929: 52)

a. ***ma*** ***gū*** *dūč-en* *nar-tõn*
LOC/DOM cow milk-PTCP.PRS can-CONT
'He can milk the/a cow.'

b. ***gū*** *dūč-en* *na* *nar-tõn*
cow milk-PTCP.PRS not can-CONT
'He cannot milk cows.'

DOM via case marking is found, e.g., in (Iron) Ossetic, illustrated in (2).

(2) Ossetic (Iron), Iranian, DOM via case marking (Abaev 1964: 142)

a. *fexst-on* ***dur***
throw:PST-1SG stone
'I threw a stone.'

b. *fexst-on* ***dur-ə***
throw:PST-1SG stone-ACC/DOM
'I threw the stone.'

Even though DOM is robustly attested for many modern Iranian languages, the two largest Iranian languages besides Persian, viz. Pashto and Kurdish, do not have DOM.

Iemmolo (2013), building upon de Hoop and Malchukov (2007: 1640, 2008), distinguishes between symmetric and asymmetric strategies in DOM (see also Chappell and Verstraete 2019). A symmetric DOM alternation involves two or more different overt markers (e.g., a genitive and an accusative case marker), whereas asymmetric DOM alternates between unmarked and marked NPs, as in examples (1) and (2) above. Example (3) shows a symmetric DOM alternation (direct object in bold once in accusative, once in genitive case) in Young Avestan.

(3) Young Avestan, Iranian, symmetric DOM (Vendidad 8.32)[2]

a. *yaṯ maśiiō maṣīm* ***xṣ̊udrå̊*** *auui fraŋhərəzaⁱti*
when man:NOM.SG man:ACC.SG semen:ACC.PL into ejaculate:IND.PRS.3SG
'when a man ejaculates (his) semen into a man.'

b. *yaṯ vā maśiiō maśiiānąm* ***xṣ̊udranąm*** *paragəᵘruuaiieⁱti*
when or man:NOM.SG man:GEN.PL semen:GEN.PL take:IND.PRS.3SG
'or when a man receives (some) men's semen.'

Turning to the final point of this section, manifestations of DOM differ in accordance with the part of speech of the direct object. The distinction between (personal) pronouns, personal names, and common nouns is of major importance in this regard, with personal names sometimes aligning with common nouns, sometimes with pronouns, as further illustrated in Section 5. In DOM systems, the choice between two different constructions for direct objects is typically restricted to common nouns, as seen in the preceding examples. Pronouns typically occur only in one of these two constructions, often receiving the same marker as the subset of specific or animate common nouns. For instance, in Munji the preposition *va* serves as DOM marker, cf. (4b). Whereas the use of this preposition only occurs with definite common noun objects (cf. non-specific *čoy* in (4a), it is obligatory with personal pronouns. The preposition and pronoun have in fact become inseparable (*va* + *ta* 'you' > *fta*), cf. (4c).

(4) Munji, Iranian, same marker for specific common nouns and pronouns (Grjunberg 1972: 457–458)

a. *parizod-in šəy, āvər* ***čoy*** *na potṣopūr-ān*
fairy-ERG go:PRET.3SG bring:PRET.3SG tea to king's.son-OBL
'The fairy went away and brought the king's son (some) tea.'

b. *potṣopūr-ān* ***va*** *yosp qāmčin kər*
king's.son-ERG OBL horse whip make:PRET.3SG
'The king's son lashed his horse.'

c. *mən* ***fta*** *ləšk'əm*
1SG.ERG 2SG.OBL see:PRET.1SG
'I saw you.'

In cases such as Munji, the same marker *va* is used for one subset of direct objects (i.e. pronouns and definite common nouns) while other direct objects remain unmarked. Importantly, however, the identity of formal marking across different parts

2 The meaning difference between the two examples relates to the specificity of the direct objects: in (3a), accusative marking on 'semen' renders this a specific direct object (indicated by the '(his)' in the translation). In (3b), the genitive marking on 'semen' allows for a non-specific or a partitive interpretation. The second genitive in this example ('men's') can be analyzed as an adnominal (possessive) genitive. The second accusative in example (3a), *maṣīm,* exemplifies a directional use of the accusative, specifying the goal of the directional preverbal adverb *auui* 'into'.

of speech is not necessary for a DOM system. The essential feature of a DOM system is that at least one subclass of nominal expressions in direct object function differs in its overt marking from other subclasses (see also Sinnemäki's 2014: 283 definition of restricted case marking and Bickel et al.'s 2015: 18–19 very detailed list of potentially relevant subclasses). With regard to marking variation across parts of speech, two options are common in addition to the one illustrated by Munji above.

The first option is to use different markers for common nouns and pronouns. Semnani, for example, exhibits an alternation between an unmarked (basic) and a marked (oblique) form of the 3rd person pronoun (e.g., *ü* 'he/she' vs. *žö* 'he.OBL'/*žin* 'she.OBL', Bossong 1985: 31). Common nouns also distinguish a basic and an oblique form, but the actual forms are quite different. Pronominal obliques make use of a former preposition (*ž-* < *až* < **hača*), while specific common noun obliques are marked with the suffix *-in*, as shown in example (5a). Non-specific common nouns remain unmarked, as shown in (5b).[3]

(5) Semnani, Iranian, different markers for nouns and pronouns

a. Specific objects, nominal and pronominal DOM

un	*ǰœnikœ-**jn***	*tâziânä*	*bukwâtœjš*	***žin***	*birin*	*kärdœjš*
that	woman-OBL	whip	beat:PRET.3SG	she:OBL	out	make:PRET.3SG

'he whipped that woman and chased her away.' (Christensen 1915: 57)

b. Non-specific object, no DOM

tâ	*xodâ*	***vačä***	*har*	*hœjr-un*	*hâ-dœj*
so.that	god	child	all	three-OBL	CONJ-give

'so that god (may) give all three (of you) a child.' (Christensen 1915: 59)

The second marking option is for common nouns to show no special marking in object function, while personal pronouns are marked in a specific way when functioning as direct objects. This appears to be the case in Semnani's close relative Lâsgerdi, where only pronouns have distinct forms when in object function, as seen in (6). The oblique form once again involves the fusion of a former preposition and a pronominal stem, but this is not crucial for the assessment of DOM. What is crucial is that pronominal direct objects behave differently from common noun direct objects.

(6) Lâsgerdi, Iranian, DOM only with pronouns

ini	***žäki***	*tâziânä*	*bökutanâš,*	***žo***	*bī*	*väkärčun*
that	woman	whip	beat:PRET.3SG	she:OBL	out	make:PRET.3SG

'he whipped that woman and chased her away.' (Christensen 1935: 81)

3 Semnani belongs to the group of Iranian languages which show "double oblique alignment" (Bickel et al. 2015: 12), where both the actor and the undergoer argument of a transitive verb may occur in the oblique case. Bickel et al. (2015: 12–14) provide a convincing account of how to accommodate these languages in the DOM variation space.

As we will show in the following sections, the kinds of morphosyntactic differences in DOM illustrated here with Iranian languages are also found in WMP languages.

2.2 A brief note on the functions of DOM

Animacy and specificity/definiteness are usually thought to be the two main dimensions relevant for the occurrence of DOM. These dimensions are typically presented in the form of implicational scales, as in I and II, where occurrence of DOM for a particular expression type implies DOM for all expression types to its left.

I. **Animacy scale**:

First/second person > third person pronoun > personal name > human > animate > inanimate

II. **Specificity/definiteness scale**:

Pronoun > personal name > definite > specific > non-specific

It appears to be the exception rather than the rule that only one dimension is relevant in a particular DOM system. In most systems, objects are differentiated with regard to both animacy AND definiteness. In the following, we simply refer to the SPECIFICITY SCALE, but this is to be understood as including both the animacy and the specificity/definiteness scales above.

Two further scales of DOM-triggering factors can be represented as in III and IV, following Bárány and Kalin (2020: 2), based on Dalrymple and Nikolaeva (2011) and Næss (2004), respectively.

III. **Information structure**:

Topic > Non-Topic

IV. **Affectedness**:

Affected > Unaffected

The exact cut-off point along the more complex scales is variable even among closely related languages and thus highly language specific, as is illustrated by the following example set in (7) from the two genetically closely related Pamir varieties Bartangi and Roshani. In both varieties, definite human arguments are regularly marked by the preposition *az*/*as*, but this also holds for non-human animates (animals) only in Bartangi.

(7) Pamir languages, Shughni-Roshani group, Iranian, DOM by means of an adposition (*az*/*as*), Sokolova (1973: 178) and Pachalina (1969: 50, 52)

a. Bartangi

az	*dim*	*aǰtur=at*	*az*	*kā*	*vūg*
DOM	this	camel=2SG	from	where	brought

b. Roshani

dum	*uǰtur=at*	*az*	*kā*	*avūg*
this	camel=2SG	from	where	brought

'From where did you bring this camel?'

Similar microvariation is observable in other languages and language families (e.g. Iemmolo 2011: 156–160 on Bantu languages, pp. 237–245 on different varieties of Italian).

The functions of DOM will not be further discussed in this paper as the sources for the languages investigated here do not provide much information in this regard. When using (IN)DEFINITE or (NON-)SPECIFIC to characterize the referential meaning of a nominal expression, this is usually based on the translation given for a particular (set of) example(s) and does not reflect a proper semantic or pragmatic investigation.

3 Are there Austronesian DOM languages?

As mentioned in the Introduction, the typological literature is sparse and inconsistent regarding instances of DOM in Austronesian languages. We limit ourselves to the four major works with more extended survey data. Bossong (1985: 3) and Iemmolo (2011: 3 and *passim*) include Austronesian among the language families that show DOM without further discussion. Bossong only mentions a few Austronesian languages in passing (Palauan and Woleaian; Bossong 1985: 116, 177), and does not provide any examples. Iemmolo (2011: 274) includes 13 Austronesian languages in his sample of DOM languages, gives examples for most of them, but does not provide an explicit argument as to why they are included. Furthermore, he does not list or discuss any Austronesian languages which, in his view, lack DOM. Sinnemäki (2014) and Bickel et al. (2015), on the other hand, work with large representative databases controlled for areal and genealogical biases which include both DOM and non-DOM languages. Sinnemäki's database comprises 744 languages, 60 of which are Austronesian, and 10 of these are considered to show asymmetric DOM (more precisely, what Sinnemäki calls "restricted case marking", as explained further shortly). Bickel et al.'s database (Witzlack-Makarevich et al. 2012) comprises 435 languages, 24 of which are Austronesian, and only one of which (namely the Oceanic language Tamambo) is considered to show DOM.

Before going into further details regarding the differences among these works, it is necessary to further limit the scope of our discussion. Austronesian is among the largest and most diverse of language families, in terms of number of languages, geographical spread and structural diversity. In typological and genealogical discussions, it is therefore sometimes divided into three major areas which allow for substantial generalizations, plus a number of "special cases". The three major areas are the Oceanic languages to the east; the Austronesian languages of the Wallacea linguistic area as defined in Schapper (2015) in the middle, covering most of eastern Indonesia;[4] and the WMP languages, spoken in western Indonesia, the Philippines, Malaysia and Madagascar, to the west (cf. Himmelmann 2005: 110–114; Adelaar and Himmelmann 2005: xvi–xvii).[5] "Special cases" are the largely isolating Austronesian languages on the Southeast Asian mainland (best known are the Chamic languages), which are heavily influenced by their non-Austronesian neighbors; the Austronesian languages of Taiwan, also known as Formosan languages, which are hypothesized to form various primary branches of the Austronesian family tree and are very diverse regarding their grammatical structures; and the two Micronesian outliers, Chamorro and Palauan. DOM phenomena have been reported for all three major areas, the Formosan languages, and Chamorro and Palauan (see Appendix 1 for details). But these phenomena are also very different in each case (in the discussion of DOM in Oceanic languages, for example, noun incorporation plays a major role: see Hopperdietzel 2020). We focus here on languages of the WMP group with symmetrical voice alternations.[6] These languages have played almost no role in typological discussions of Austronesian DOM phenomena and pose special challenges regarding their classification in typological databases, as further explained in the following section.

Table 1 lists the WMP symmetrical voice languages that have been included in the major typological DOM surveys mentioned at the beginning of this section (Bossong 1985 does not mention any of them). Recall that Iemmolo (2011) only lists

4 This area is roughly co-extensive with the Central Malayo-Polynesian (CMP) and South Halmahera-West New Guinea branches of the Austronesian family tree, both branches being somewhat controversial (Adelaar 2005b: 24–26; Zobel forthcoming).

5 While the Austronesian family tree is often represented as including a WMP branch, WMP languages are generally not considered to form a single branch of Malayo-Polynesian (MP). Rather "Western Malayo Polynesian" here and elsewhere refers to a loosely related group of languages that may contain several primary branches of MP. See Smith (2017) for a recent summary of the state of the art.

6 This includes most Austronesian languages of western Indonesia, Malaysia, the Philippines and Madagascar, with the exception of the languages of the southern half of Sulawesi (e.g. Buginese, Selayarese), the Barrier Island Languages (e.g. Nias), and Acehnese. The status of the South Sulawesi languages (Buginese, etc.) is controversial. See Himmelmann (2005: 111–114) and Kroeger and Riesberg (2024) for further details.

Table 1: Western-Malayo-Polynesian symmetrical voice languages mentioned in previous typological DOM surveys; classification as having (+) or not having (–) DOM according to the compilers of the databases.

Language	Iemmolo (2011)	Sinnemäki (2014) [non-pronominal asymmetric DOM only]	Witzlack-Makarevich et al. (2012)
Hiligaynon (hili1240)	+		
Malagasy (mala1537)	+	+	
Tagalog (taga1270)	+	–	
Begak-Ida'an (idaa1241)	+	–	
Manadonese (mala1481)	+	+	
Tboli (tbol1240)		+	
Limos Kalinga (limo1248)		–	
Kapampangan (pamp1243)		–	
Tagabawa (taga1272)		–	
Indonesian (indo1316)		–	
Riau Indonesian (riau1236)		–	
Minangkabau (mina1268)		–	
Sundanese (sund1252)		–	
Batak (Toba; bata1301)			–
Balangao (bala1310)			–
Bontok (bont1247)			–
Gorontalo (goro1259)			–
Ilocano (ilok1237)			–
Javanese (java1254)			–
Madurese (nucl1460)			–
Mualang (mual1241)			–

languages with DOM, while the samples by Sinnemäki (2014) and Witzlack-Makarevich et al. (2012) include DOM and non-DOM languages.

Table 1 lists 21 languages of which 17 are classified as not showing DOM by either Sinnemäki or Witzlack-Makarevich et al. For nine of these 17 languages, we will argue in Section 5 that they do in fact show DOM. For two languages – Tagalog and Begak-Ida'an – we find diverging classifications in the table: they are +DOM in Iemmolo (2011), but –DOM in Sinnemäki (2014). The obvious question that arises from Table 1 is why WMP symmetrical voice languages tend to be classified as non-DOM languages, and why different investigators arrive at different classifications.

There are several reasons for the fate of WMP symmetrical voices languages in cross-linguistic work on DOM phenomena. To begin with an almost trivial one, grammatical descriptions of these languages rarely discuss DOM phenomena and, if they do, they do not discuss it under the label of DIFFERENTIAL OBJECT MARKING. There is also hardly any specialist work dedicated to DOM in a WMP language, Latrouite (2011) on Tagalog being the major exception.

Turning to more substantial reasons, different classifications also result from differences in the definitions used in compiling a database. Sinnemäki, for example, limits his survey to full noun phrases, excluding DOM phenomena which are limited to personal pronouns (Sinnemäki 2014: 287). Bickel et al. (2015: 18), on the other hand, have an extensive coding scheme for pronouns. Furthermore, Sinnemäki (2014: 286) excludes symmetric DOM, limiting himself to what he calls restricted case marking of direct objects, i.e. overt case marking that is limited to a subset of objects, other objects remaining (morphologically) unmarked (2014: 283). Bickel et al. (2015: 10–15), on the other hand, define differential argument marking[7] in terms of marked and unmarked alignment sets, where overt morphological marking becomes only indirectly relevant. The major criterion is how the morphosyntactic markings of the syntactic functions A, P, and S relate to each other with respect to a particular set of constructions (e.g., nominal arguments in simple transitive and intransitive clauses). If all functions receive the same kind of marking, they form the alignment set {S = A = P}, which is more general and less marked than {S = P} or {S = A}, which in turn are less marked than {A}, {P} or {S} (cf. also Schmidtke-Bode and Levshina 2018: 511–512 for illustration and discussion). All WMP symmetrical voice languages in their sample (Witzlack-Makarevich et al. 2012) are coded as making use of the most general alignment set {S = A = P} in all investigated constructions, and consequently as lacking differential A and P marking.

This is not the place to further evaluate the advantages and disadvantages of the DOM definitions used in Sinnemäki (2014) and Bickel et al. (2015), as this, to our minds, is not the main reason for the problematic classification of WMP symmetrical voice languages in DOM databases. The main reason is the fact that applying whatever the criteria for DOM are (including the ones used by Sinnemäki and Bickel et al.) presupposes an analysis of how these criteria play out in a symmetrical voice system, an extra step in compiling a typological database, which is not straightforward and often not taken. The next section is dedicated to this task before we then reevaluate the WMP symmetrical voice languages listed in Table 1 regarding their DOM status in Section 5.

4 WMP symmetrical voice languages and DOM: preliminaries

Symmetrical voice systems are typologically rare and make the identification of DOM less straightforward than it might be in better-known DOM languages. This section

7 Bickel et al. (2015) investigate differential agent/subject marking as well as differential patient/object marking. Here, we are only concerned with their results pertaining to differential patient marking.

therefore first introduces WMP symmetrical voice systems (4.1), and then defines (4.2) what we do, and do not, include as instances of DOM in our survey of the phenomenon in WMP symmetrical voice languages. Importantly, the peculiarities of the symmetrical systems themselves have led to a bewildering number of different analyses and idiosyncratic terminological choices which, we submit, constitute major obstacles when determining DOM phenomena within this group of languages.

4.1 Symmetrical voice

Voice is generally considered a mechanism that selects one semantic argument over others to become the privileged syntactic argument (PSA or subject) of the clause (Shibatani 1988; Zúñiga and Kittilä 2019, among many others). In most of the world's languages, voice alternations go hand in hand with a decrease in transitivity. That is, the alternation between an active and a passive construction involves the demotion of the actor argument, and the alternation between an ergative and an antipassive construction involves the demotion of the undergoer argument. Symmetrical voice alternations as found in many WMP languages (cf. Foley 1998, 2008; Himmelmann 2005; Riesberg 2014), on the other hand, are alternations that link different semantic arguments to subject function without demoting the non-subject argument. Consequently, these languages exhibit two (or more) basic transitive constructions which are usually called ACTOR VOICE (AV) and UNDERGOER VOICE(S) (UV) in the current literature. In AV, the subject of the construction bears the agent role, and the verb will be marked by AV morphology. In UV, the subject is an undergoer-like argument (patient, theme, etc.), which in turn will be signaled by UV morphology on the verb. Example (8) illustrates this for Totoli, a WMP language of Sulawesi, Indonesia. In (8a) the subject (the third person singular pronoun *isia*) is the agent, and the verb carries the AV (realis) prefix *noN-*. In (8b), the verb is marked by the UV (realis) prefix *ni-*, and the subject of the construction is the undergoer argument *saginna* 'his/her banana'.[8] In both examples, subject arguments and voice morphology are marked in bold in glosses and translation. Importantly, the non-subject arguments (*saginna* 'her/his banana' in (8a) and *kami* 'we' in (8b)) remain direct core arguments, which are generally restricted to immediate postverbal position, and both constructions are fully transitive.

8 The fact that the two constructions in (8) display two different word orders – SVO in AV and VOS in UV – does not constitute a difference between the AV and UV constructions in general. Both voice constructions allow for both orders, i.e. also VOS in AV, and also SVO in UV. However, the data in (8) reflect a discourse preference in Totoli, in that AVs with SVO order and UVs with VOS order are much more frequent than the respective other option (cf. Riesberg et al. 2019: 537).

(8) Totoli (elicited)

a. *Isia nangaan saginna.*
isia **no**N-kaan sagin=na
3SG **AV.RLS**-eat banana=3SG.GEN
'**S/he** ate her/his banana.'

b. *Nikaanmo kami saginna.*
ni-kaan=mo kami **sagin=na**
UV.RLS-eat=CPL 1PL.EXCL **banana=3SG.GEN**
'We ate **her/his banana**.'

The languages of the Philippines and adjacent areas of Indonesia usually exhibit two or three semantically distinct UVs that differentiate more fine-grained semantic roles of their subject arguments. Example (9) shows a symmetrical alternation between an AV construction in (9a) and three UV constructions in (9b–d) in Tagalog (again, subject arguments and the voice morphology are highlighted in bold). The three UVs differ in whether their subject is a PATIENT/THEME in (9b), a LOCATIVE in (9c), or a BENEFACTIVE in (9d). For languages like Tagalog, it is thus common practice to distinguish between PATIENT VOICE (PV), LOCATIVE VOICE (LV), and CONVEYANCE VOICE (CV) which are all subsumed under the more general umbrella term UNDERGOER VOICE.

(9) Tagalog (Foley and van Valin 1984: 135)[9]

a. Actor voice
*B<**um**>ili* ***ang*** ***lalake*** *ng* *isda* *sa* *tindahan.*
<**AV.RLS**>buy **NOM** **man** GEN fish LOC store
'**The man** bought fish at the store.'

b. Patient voice
*B<in>ili-**Ø*** *ng* *lalake* ***ang*** ***isda*** *sa* *tindahan.*
<RLS>buy-**PV** GEN man **NOM** **fish** LOC store
'The man bought **the fish** at the store.'

c. Locative voice
*B<in>ilh-**an*** *ng* *lalake* *ng* *isda* ***ang*** ***tindahan***.
<RLS>buy-**LV** GEN man GEN fish **NOM** **store**
'The man bought fish **at the store**.'

d. Conveyance voice
I**-b-in-ili* *ng* *lalake* *ng* *isda* ***ang ***bata***.
CV-<RLS>buy GEN man GEN fish **NOM** **child**
'The man bought fish **for the child**.'

9 Glosses in quoted examples do not always correspond to the glosses used in the original sources but have been adapted for reasons of consistency. Accentual diacritics, which in some uses indicate vowel length, in others 'stress', are generally omitted. Glottal stops are indicated by <'>.

As the comparison of the Totoli and Tagalog examples reveals, the two languages do not only differ in their number of UVs. They also exhibit different argument realization strategies. In Totoli, core arguments are unmarked, whereas in Tagalog every noun in core function is preceded by a case marker (i.e., nominative *ang*, genitive *ng*, or dative *sa* in (9)). There is thus considerable variation regarding the formal and functional properties of the symmetrical voice systems attested in WMP languages. In the specialist literature, it is common practice to distinguish two basic types of WMP symmetrical voice languages. The "Philippine-type" languages behave like Tagalog and are characterized, *inter alia*, by more than three symmetrical voice alternations and overt case marking proclitics. The "non-Philippine-type" languages typically only have two productive voice alternations and lack overt case marking. The latter group is essentially negatively defined, reflecting the fact that it is much more heterogeneous than the former one. Non-Philippine-type languages have also been called "Indonesian-type languages", but this term has recently been used in a number of different senses, so that "non-Philippine-type" is more precise (Himmelmann 2005: 111–114; McDonnell and Chen 2022: 13). The remainder of this section deals only with Philippine-type languages, as they provide the clearest evidence for those properties of symmetrical voice systems that impact the analysis of DOM.

In the preceding exposition of the basic features of symmetrical alternations we use the term 'subject' which, however, has been the object of a longstanding controversy in Philippine linguistics. In his influential paper, Schachter (1976) summarizes a debate that had already started in the late 1950s and asks, whether the Philippine subject is "topic, actor, actor-topic, or none of the above?" One common view at the time was that the nominative phrase constitutes the topic of the clause, while the actor phrase is always to be analyzed as subject. Under such an approach, alternations like in (9) do not involve a change of grammatical relations and consequently were not considered voice alternations. Instead, the term "focus" was adapted (see Blust 2002: 64) and constructions like (9a–d) were referred to as "agent focus", "patient focus", etc. Other terms used in the Austronesian literature are (verbal) "case", "topicalization", and "trigger" (see Blust 2002: 73–74 for an overview). Today, mainly following the arguments presented by Kroeger (1993a, 1993b), the common view in the Austronesian specialist literature is that the nominative marked argument constitutes the privileged syntactic argument and that it can indeed be called "subject" (Chen and McDonnell 2019: 117–118; Himmelmann 2005: 152–157; Kroeger and Riesberg 2024).

The bewildering diversity of analytical and terminological proposals does not only pertain to the nature of the voice alternations and the verbal morphology. It also extends to nominal marking. The terminological variation in the literature here is even worse than that for the voice alternations. According to Blust (2015: 437–39), there are (at least) 42 different terms for what we call "case markers" here, including

"articles", "attributive and object markers", "auxiliary nouns", "introducers", "noun class markers", "orienters", "specifying nouns", or "topic/goal markers". It should be obvious that such heterogeneity makes for a major stumbling block when trying to include data from WMP symmetrical voice languages in a typological database. Non-experts hardly stand a chance of being able to distinguish between substantial grammatical differences and more superficial terminological differences. In the present work, we follow the currently most widely used terminology and speak of "case markers" (see Himmelmann 2021 for further discussion).

4.2 What is, and what is not, DOM in WMP symmetrical voice languages

In the previous section, it was shown that symmetrical voice alternations *per se* pose a number of analytical and conceptual challenges which have resulted in several competing analyses and terminologies which, in turn, make it difficult to identify related constructions across different languages. Here we address the issues that specifically pertain to the identification of DOM constructions.

When investigating DOM in languages that exhibit more than one basic transitive construction, the question arises which of these should be included in the analysis. None of the studies we scrutinize in this paper explicitly state which basic transitive construction(s) entered their typological sample. Neither Sinnemäki (2014) nor Witzlack-Makarevich et al. (2012) provide any information on this topic. Iemmolo (2011), too, does not explicitly state which constructions type(s) he included in his study. The examples from symmetrical voice languages he presents, however, are exclusively AV constructions.

In our investigation we only consider AV constructions to involve DOM for the following reasons. First, in UV the more patient-like argument occurs in subject function but DOM pertains to objects (i.e. core arguments that are not in subject function). Second, in AV the more patient-like argument fulfills the definition for direct objects given in Section 2, i.e., it is the more patient-like core argument of a prototypical two-place transitive predicate, the morphosyntactic marking of which differs from the marking of the single core argument of an intransitive predicate (which always occurs in the nominative *ang*-form). Third, there are different marking options for direct objects in AV. To illustrate this, we first have to take a closer look at case marking.

Case markers in Philippine-type languages usually come in two paradigms – one for personal names and one for common nouns – and they often distinguish three different cases, usually labelled "nominative", "genitive" and "dative/locative". Additionally, personal pronouns often exhibit distinct forms for each case. Table 2

Table 2: Case markers and 1SG pronoun forms in Tagalog.

	NOM	GEN	DAT/LOC
Common noun markers	*ang*	*ng*[a]	*sa*
Personal name markers (singular)[b]	*si*	*ni*	*kay*
Personal pronouns (1SG)	*ako*	*ko*	(*sa*) *akin*

[a]The case marker is /naŋ/ which, however, is conventionally written as ⟨ng⟩. [b]Personal name markers often have a special form for associative plurals. Thus, Tagalog also has *sina*, *nina*, and *kina* for associative plural personal names. These plural forms are not included in the tables in Section 5.

shows the basic case markers and the three case forms of the first person singular in Tagalog.

The default pattern of case marking and voice alternations of a Philippine-type symmetrical voice language is as follows: Each verbal clause[10] must contain one nominative argument and a voice marker on the verb marking the semantic role of the nominative argument in the clause (cf. the examples in (9) above). All other NP constituents of the clause are formally marked as genitive or dative/locative case, depending on their semantic role: genitive *ng* occurs with actors and experiencers in UV, and with patients, themes, and instruments in AV; dative/locative *sa* occurs with goals, recipients, locations, as well as with temporal adjuncts. The dative personal name marker *kay* is mostly used for recipients (and DOM).

As we will see in Section 5, both the formal and the functional properties of the case marking system are variable across Philippine-type languages, sometimes with important implications for DOM. While our sources use a wide range of labels for the case forms, we will generally speak of nominative, genitive and dative case forms, as just defined. In addition, we will use the label OBLIQUE for case forms that combine genitive and dative functions and thus occur with the full range of non-subject arguments, including agents, patients, themes, instruments, recipients, goals, and locations.

The different case forms for common nouns, personal names, and pronouns, as illustrated in Table 2 for Tagalog, do not *per se* constitute an instance of differential argument marking. That is, the mere fact that case marking for common nouns takes another form than case marking for personal names (e.g., genitive *ng* for common nouns vs. genitive *ni* for personal names) is not DOM. What does count as DOM, however, are those instances where nominal expressions of different classes are marked by different cases in the same grammatical context. This is illustrated in (10),

10 As discussed in Himmelmann (2008) and Kaufman (2009), *inter alia*, the distinction between verbs and nouns in Philippine-type languages is not as straightforward as in other language families. Here, we follow the widespread usage of calling voice-marked forms "verbs".

where the common noun in (10a) receives genitive case, and the personal name in (10b) receives dative case.

(10) Tagalog (Latrouite 2011: 94–95)

a. *Siya* *ang* *naka-kita* ***ng*** *aksidente.*
3SG.NOM NOM POT.AV.RLS-visible GEN accident
'He is the one who saw a/the accident.'

b. *Siya* *ang* *naka-kita* ***kay*** *Jose.*
3SG.NOM NOM POT.AV.RLS-visible DAT Jose
'He is the one who saw Jose.'

Our definition of DOM, of course, also includes cases where a given noun class allows for alternating case markers in the same construction, as illustrated in (11). Here, both sentences contain the non-subject undergoer argument, *daga'* 'cat', but this is either marked by genitive *ng*, as in (11a), or by dative *sa*, as in (11b). As reflected in the English translation, the differential use of the dative (instead of the genitive) evokes a difference in definiteness, in that dative marked phrases are unambiguously definite, while the genitive marked undergoer can be understood as either definite or indefinite (specific).

(11) Tagalog (Himmelmann 2005: 148; McFarland 1978: 157)

a. *Ito* *ang* *pusa-ng* *k<um>ain* ***ng*** ***daga'.***
PRX NOM cat-LK <AV>eat GEN **rat**
'This is the cat that ate a/the rat.'

b. *Ito* *ang* *pusa-ng* *k<um>ain* ***sa*** ***daga'.***
PRX NOM cat-LK <AV>eat DAT **rat**
'This is the cat that ate the rat.'

In concluding this section, a final note on the possibility of DOM in undergoer voice constructions is in order. Strictly speaking – taking the symmetry of the system seriously – actors in UV constructions such as (8b) and (9b–d) are direct objects in that they are non-subject core arguments of a transitive predicate marked differently from the single core argument of an intransitive predicate. But they are not the patient-like core argument, and hence any variation in the marking of the agent argument in object function (which in fact occurs in some WMP languages) would not count as DOM.

5 DOM in WMP symmetrical voice languages: survey

In this section, we report on our survey of DOM in WMP symmetrical voice languages. Our main concern is the 21 languages listed in Table 1, which have been

claimed to have, or lack, DOM in the previous literature. Table 5 in Section 5.4 summarizes our findings regarding these languages. In addition, we surveyed all WMP symmetrical voice languages[11] for which sufficiently substantial grammatical descriptions were available to us, in order to provide for a more complete picture of the distribution of DOM phenomena in these languages. All in all, data from 57 WMP symmetrical voice languages were scrutinized for evidence of DOM, which was found for 34 of them. Appendix 1 provides the details of this survey, including language names, subgroup affiliation, sources consulted, and type of DOM, if applicable.

The sources consulted for the survey are written in a wide variety of theoretical frameworks and very often lack explicit information on DOM. The basic notion of differential object marking is absent from nearly all of them. Apart from a few works that clearly state and illustrate the occurrence of DOM (often not using the term, though), even fewer works explicitly deny the occurrence of DOM. This means that evidence of DOM had to be searched for in sections on case marking paradigms (recall the bewildering variety of labels for these from Section 4.1), AV constructions, relative clauses and cleft constructions, the latter two being the constructional contexts most likely to show specific undergoers in non-subject function. We also searched for specific forms or glosses (e.g., elements glossed as 'oblique' or 'dative') in the examples used in the grammars and in the texts occasionally provided by the source. Given these limitations of our sources, it is very likely that the current survey under-reports DOM in WMP, and the judgment "no DOM" strictly means "no evidence for DOM identified in the sources used for this study".

Our search for evidence of DOM was focused on the three top levels of the specificity scale, another important consequence of the limitations of our sources. That is, we looked for evidence of DOM specifically for (personal) pronouns, personal names and common nouns, as these are the empirical domains most likely to be exemplified in our sources. Note that personal names often receive specific marking in WMP languages (e.g., honorific articles) and therefore are regularly mentioned even in shorter grammatical sketches. Most sources do not address animacy, information structure or affectedness as potentially relevant factors for

11 Recall the delimitation of this group of languages given in Section 3 above. Importantly, our sample does not include Formosan languages and languages from southern Sulawesi such as Wolio, Muna, Selayarese, Makassarese or Buginese, some of which have been claimed to be symmetrical voice languages. We likewise leave aside Chamorro and Palauan as their grammatical systems are substantially influenced by their heavy contact with Spanish (in the case of Chamorro), and with Spanish, German, Japanese and English in the case of Palauan (Josephs 1984). Furthermore, while both languages have been analyzed as symmetrical voice languages, the precise nature of the voice alterations in both languages still needs further scrutiny.

DOM (cf. Section 2.2 above). Regarding specificity/definiteness, there is usually no information as to the more precise parameters that play a role; for example, is definiteness (excluding specific indefinites) or is specificity (including specific indefinites) relevant?

Despite these limitations, our focus on pronouns, personal names and common nouns allows us to ask whether WMP symmetrical voice languages provide evidence for the specificity scale, thereby supporting the universalist view on DOM. Furthermore, the specificity scale provides us with a structure for presenting our data. In accordance with the scale, we distinguish three basic types of WMP languages with regard to DOM. Type I languages show evidence for DOM at all three levels: pronouns, personal names, and common nouns. Type II languages show evidence only with regard to personal pronouns and names and Type III languages restrict DOM to (singular) personal pronouns.

5.1 Type I: DOM for pronouns, personal names, and common nouns

Type I DOM is particularly common in Philippine-type languages. Since in general all common nouns and personal names in these languages are overtly case marked, Type I DOM is usually of the symmetric type, as illustrated in Section 5.1.1. However, there are, much more infrequently, also Type I languages of both the Philippine- and the non-Philippine-type, where nominal expressions are not overtly case marked in all grammatical functions and where, consequently, we find asymmetric manifestations of DOM, as illustrated in 5.1.2. For further details on variation and complications in determining DOM in Philippine-type languages, Appendix 2 discusses the Philippine language Cebuano.

5.1.1 Standard symmetric examples of Type I DOM: Tagalog, Hiligaynon and Pangasinan

DOM in Tagalog pertains to the alternation of genitive and dative proclitics when marking non-subject undergoers in AV constructions (cf. Section 4.2 above). Examples (10) and (11) in Section 4.2 already provided partial exemplification. In (12) we provide the full example set for common nouns (a, b), personal names (c) and personal pronouns (d).

(12) Tagalog, symmetric DOM with common nouns, pronouns, and personal names (Latrouite 2011: 94–96).

a. *Siya* *ang* *naka-kita* ***ng*** *aksidente*
3SG.NOM NOM POT.AV.RLS-visible GEN accident
'He is the one who saw a/the accident.'

b. *Siya* *ang* *naka-kita* ***sa*** *aksidente*
3SG.NOM NOM POT.AV.RLS-visible DAT accident
'He is the one who saw the accident.'

c. *Siya* *ang* *naka-kita* ***kay*** *Jose*
3SG.NOM NOM POT.AV.RLS-visible DAT Jose
'He is the one who saw Jose.'

d. *Siya* *ang* *naka-kita* ***sa*** *akin*
3SG.NOM NOM POT.AV.RLS-visible DAT 1SG.DAT
'He is the one who saw me.'

e. *Siya* *ang* *naka-kita* ***ng/*sa*** *kaniya-ng* *asawa*
3SG.NOM NOM POT.AV.RLS-visible GEN/DAT 3SG.DAT-LK spouse
'He/She is the one who saw his/her spouse.'

Dative *sa* optionally marks specific common nouns, whereas pronouns and personal names in non-subject undergoer function are always in the dative form. In the case of pronouns, this also involves the dative proclitic *sa*, while personal names have their own dative marker *kay*. All nominal expressions are overtly case marked in all syntactic functions (cf. Table 2 above), hence this is an instance of symmetric DOM.

Latrouite (2011) further notes that inherently definite common noun expressions such as nouns modified by a possessor phrase (cf. 'his/her spouse' in (12e)) generally do not allow dative marking, despite being definite and specific. We mention this here to make it clear that there is more to the factors conditioning DOM in WMP symmetrical voice languages than a simple specificity distinction, specificity itself being a multifaceted notion. Since we simply want to demonstrate the occurrence of DOM in these languages, we refrain from discussing such details for reasons of space and lack of relevant data for most of the languages in our sample.

DOM in Hiligaynon is also symmetric, showing the same basic pattern as Tagalog with slightly different case forms. As illustrated in (13a), indefinite non-subject arguments in AV are marked by genitive *sang* (parallel to Tagalog *ng*), while definite non-subject core arguments – just like goals and recipients – take the dative marker *sa* (compare (13a) and (13b)).

(13) Hiligaynon

a. *Nag-hatag* *ang* *maestra* ***sang*** *bola* *sa* *bata'*
AV.RLS-give NOM teacher GEN ball DAT child
'The teacher gave a ball to the child.' (Spitz 2002: 386)

b. *Nag-hakos ako **sa** propesor.*
AV.RLS-hug 1SG.NOM **DAT** professor
'I hugged the professor.' (Spitz 2001: 13)

While in the case of common nouns the choice of genitive *sang* versus dative *sa* covaries with definiteness (and possibly other factors), pronouns and personal names always take dative marking, as illustrated in (14).

(14) Hiligaynon, DOM with pronouns and personal names
a. *Nag-hampa' ang maestra **sa** iya*
AV.RLS-strike NOM teacher **DAT** 3SG.UNF
'The teacher hit him/her.' (Spitz 2002: 385)
b. *Nag-hampa' ang maestra **kay** Roberto*
AV.RLS-strike NOM teacher **DAT** Roberto
'The teacher hit Roberto.' (Spitz 2002: 386)

A third example for standard Type I DOM in Philippine-type languages is Pangasinan. As seen in (15), there is again minor variation of the case forms across different parts of speech.

(15) Pangasinan, DOM in common nouns (Benton 1971: 193)[12]
a. *Siak so angan **na** mansanas.*
1SG NOM AV:eat **GEN** apple
'I ate a/the apple.'
b. *Samay manok so angan **ed** mansanas.*
DIST chicken NOM AV:eat **DAT** apple
'The chicken ate (of) the apple'
c. *Mangibangat si Juan **kinen** Maria*
AV:teach NOM Juan **DAT** Maria
'John/Juan will teach Maria.'
d. *Mangibangat si Juan **ed** siak*
AV:teach NOM Juan **DAT** 1SG
'John/Juan will teach me.'

There are two minor differences between Pangasinan and languages like Tagalog and Hiligaynon. First, Pangasinan does not distinguish a dative case series for personal pronouns. Instead, personal pronouns are marked with the same case markers as common nouns (compare *ed mansanas* 'DAT apple' in (15b) with *ed siak* 'DAT 1SG' in (15d), where *siak* is the full or independent form of the pronoun which also occurs in topic and predicate functions). This is different from Tagalog, where *sa* occurs side by

12 Translations modified in line with Benton's explanations.

Table 3: DOM-related case marking in Tagalog, Hiligaynon, and Pangasinan.

	Tagalog			Hiligaynon			Pangasinan		
	Gen	Dat	DOM	Gen	Dat	DOM	Gen	Dat	DOM
Common nouns	*ng*	*sa*	*ng* *sa*	*sang*	*sa*	*sang* *sa*	*na*	*ed*	*na* *ed*
Personal names	*ni*	*kay*	*kay*	*ni*	*kay*	*kay*	*nen*	*kinen/ed si*	*kinen/ed si*
Pronouns (1sg)	*ko*	*sa akin*	*sa akin*	*ko*	*sa akon*	*sa akon*	*ko*	*ed*	*ed siak*

side with the obl pronoun set. Second, for personal nouns, dative can be marked by a special form of the personal name marker, i.e. *kinen* as in (15c), or it can also be expressed with dative *ed* (hence *ed si Maria* would also be possible). In both instances, we find variability in the marking of dative case forms, which in fact is frequently found in Type I languages, as further illustrated with Cebuano data in Appendix 2.

Table 3 summarizes the forms and functions of the case markers relevant for DOM in Tagalog, Hiligaynon, and Pangasinan. This table serves as the basis of comparison for the further variants of DOM reviewed in the next sections and Appendix 2. The important point to take note of for now is that Type I DOM always involves dative or dative-like marking for specific non-subject undergoer arguments, while other non-subject arguments receive other types of marking. In standard Type I DOM, the "other type of marking" is overt genitive marking, but there are other options, as we will see in the next section.

There are 14 languages with Type I DOM in our sample, as documented in Appendix 1. Most of these languages show the same distributions and marking options illustrated in Table 3. But there are also cases that can be considered to belong to the same basic type, but with the important difference that they involve asymmetric DOM, which is of particular relevance in the current context, because Sinnemäki's (2014) database only includes (non-pronominal) asymmetric DOM systems.

5.1.2 Asymmetric variants of Type I: Malagasy, Manado Malay, Kapampangan

The clearest example of an asymmetric Type I DOM system is the one found in Malagasy (also classified as such by Sinnemäki 2014). Indefinite[13] non-subject undergoers generally remain unmarked as in example (16a), while definite non-

13 We follow here the literature where the Malagasy system is usually discussed in terms of definiteness (Howe 2022; Iemmolo 2011: 205–207; Keenan 2008; Pearson 2001; Zribi-Hertz and Mbolatianavalona 1999).

subject undergoers are often marked with the dative marking proclitic *an* in addition to a determiner, as illustrated in (16b, c).

(16) Malagasy, asymmetric Type I DOM
  a. *Manao* ***(*an)*** ***farafra*** *mahafinaritra* *io* *mpandrafirtra* *io*
  PRS.AV:make **bed** pleasing DET carpenter DEM
  'That carpenter makes pleasing beds.' (Keenan 2008: 246)
  b. *Mijery* ***an*** ***ilay*** ***alika*** *Rasoa*
  PRS.AV:watch **DAT** **DET** **dog** Rasoa
  'Rasoa is watching this dog.' (Zribi-Hertz and Mbolatianavalona 1999: 191)
  c. *Tsy* *mahalala* ***an*** ***io*** ***olona*** ***io*** *aho*
  NEG PRS.AV:know **DAT** **DET** **person** **DEM** 1SG
  'I don't know that person.' (Iemmolo 2011: 206)

Dative marking is obligatory for personal nouns and pronouns, as illustrated in (17). Malagasy object pronouns have a fused reflex of the dative marker *an* (Adelaar and Kikusawa 2014: 502–504), a phenomenon typical for DOM systems, as briefly discussed in Section 2.1 above.

(17) Malagasy, asymmetric Type I DOM (Keenan 2008: 245)
  a. *Nanenjika* ***ahy*** *izy*
  PST.AV.chase 1SG.DAT 3SG
  'He chased me.'
  b. *Nanenjika* ***an-dRabe*** *aho*
  PST.AV.chase DAT-Rabe 1SG
  'I chased Rabe.'

Malagasy differs from the Philippine-type languages discussed in the preceding section in that there is no overt case marking for core arguments. This is also the case in Manado Malay, a Trade Malay variety (Adelaar 2005a) spoken in eastern Indonesia. Pronouns and personal names in non-subject undergoer function here are generally marked with the directional preposition *pa* (18a, b), which can also optionally be used for specific non-subject undergoers (18c).

(18) Manadonese (Manado Malay), asymmetric Type I DOM (Iemmolo 2011: 104)
  a. *Utu* *da* *skop* ***pa*** *kita*
  Utu PST kick **DIR** 1SG
  'Utu kicked me.'
  b. *Kita* *da* *tampeleng* ***pa*** *John*
  1SG PST slap **DIR** John
  'I slapped John.'

c. *Dorang da kuti **(pa)** itu anak*
3PL PST flick **DIR** the child
'They flicked the child.'

A similar distribution is found in Baba Malay (Lee 2022: 134–135), where the relevant preposition is *sama* 'with, together'. It is likely that most other Trade Malay varieties also show these constructions though the available sources often do not provide the relevant information (see Bossong 2021 for some pointers). Note that Trade Malay varieties (as well as the major Standard Malay varieties such as Indonesian) typically have much larger inventories of prepositions than typical Philippine-type languages, often of recent origin. It is thus not unlikely that the DOM constructions found in these varieties involve a different developmental trajectory than the ones attested in Philippine-type languages.

The DOM system attested in Kapampangan (or Pampangan) brings us back to the Philippine-type languages discussed in the preceding section. Kapampangan differs from these in several important aspects, one of them being the (a)symmetry of the system. In Kapampangan, genitive case is restricted to non-subject actors and possessors and thus is not used with non-subject undergoers. Instead, non-subject undergoers are generally marked with the dative case, as illustrated for specific common nouns in (19a), personal names in (19b) and pronouns in (19c).

(19) Kapampangan, asymmetric Type I DOM
a. *menaya=ya **king anak** ing lalaki*
PST.AV:wait=3SG.NOM **DAT** **child** NOM man
'The man waited for a/the child.' (Richards 1971: 119)
b. *Dinatang ne ing ipus a sumaup **kang Ara***
AV:arrived COMP:3SG.NOM NOM servant LK AV:will.help **DAT** **PN**
'The servant arrived who was going to help Ara.' (Mithun 1994: 259–260)
c. *Ninung mikpuk **keka**?*
who:LK hit:AV **2.DAT**
'Who hit you?' (Mithun 1994: 262)

As seen in these examples, Kapampangan also has a system of obligatory clitics indexing core arguments in the verb complex. In AV, it is only the subject that is obligatorily cross-referenced on the predicate of main clauses (in example (19a) *ya* cross-references *ing lalaki*; in (19b) *ya* occurs in a portmanteau form with the completive particle *na* (*na* + *ya* =*ne*)).

Returning to DOM, the important difference between Kapampangan and the Philippine-type languages in Section 5.1.1 is that genitive case cannot be used for non-specific undergoers in AV constructions. Rather, non-specific undergoers in AV remain either unmarked for case, as in (20a), or they are linked to the verb complex by a linking element that otherwise appears in various types of modification

Table 4: DOM-related case marking in Tagalog and Kapampangan.

	Tagalog			Kapampangan		
	Gen	Dat	DOM	Gen	Dat	DOM
Common nouns	*ng*	*sa*	*ng* *sa*	*ning*	*king*	*Ø/-ng* *king*
Personal names	*ni*	*kay*	*kay*	*=ng*	*kang*	*kang*
Pronouns (1sg)	*ko*	*sa akin*	*sa akin*	*=ku*	*kanaku/kaku*	*kanaku/kaku*

constructions (e.g., adjective and noun, adverbial modifiers), as in (20b). Specific undergoers in AV, on the contrary, are marked with the dative case marker *king*, as shown in (20c).

(20) Kapampangan, asymmetric Type I DOM (Richards 1971: 119–121)

a. *ing lalaki ing menakit* ***anak***
NOM man NOM PST.AV.see **child**
'The man is the one who saw a child.'

b. *menaya=ya-**ng*** ***anak*** *ing lalaki*
PST.AV.wait=3SG.NOM-LK **child** NOM man
'The man waited for a child (non-specific).'

c. *ing lalaki ing menakit* ***king*** ***anak***
NOM man NOM PST.AV.see DAT **child**
'The man is the one who saw the child.'

Thus, with non-specific undergoers in non-subject function, a special construction is used which does not involve an overt case marker but a bare noun. Hence, the Kapampangan system can be argued to be an instance of asymmetric DOM despite the fact that common noun expressions are overtly case marked in most argument functions.

Table 4 provides a comparison of the Tagalog and Kapampangan DOM systems. The basic layout of the systems is the same. The major difference pertains to the use of the genitive and dative forms, with Kapampangan not allowing the genitive to be used for non-subject undergoer functions. Instead, non-subject undergoers are generally marked by the dative case marker or, if non-specific, they remain unmarked (= zero marking or linker).

5.2 Type II: DOM in pronouns and personal nouns only (Tboli, Toratán)

In a few languages in our sample (seven in all), DOM only occurs with personal names and pronouns. We briefly discuss two examples, which together give an idea of the range of variation that occurs among Type II languages.

In Tboli, common nouns occur in the unmarked base form in all core argument functions, including non-subject undergoer, as illustrated in (21a) for a non-specific undergoer and in (21b) for a specific undergoer. The latter often, but not obligatorily, are preceded by a demonstrative. Personal names in non-subject undergoer function are preceded by the oblique marker *ke*, as seen in (21c). Personal pronouns also occur in a special oblique form when in non-subject undergoer function (21d) and also *kut* in (21a)).

(21) Tboli, DOM in personal nouns (Forsberg 1992: 66–67)

a. *Mangay-i* *kut* ***el***
AV:fetch-2SG.NOM 1DU.OBL water
'Please fetch some water for us.'

b. *Omin* *le* *mtem* ***yem*** ***gunù***
then 3PL.NOM AV:burn that house
'And then they burnt that house.'

c. *Ne* *omin* *le* *t<m>udà* ***ke*** *Was*
and then 3PL.NOM <AV>throw OBL Monkey
'And then they threw Monkey out.'

d. *Du* *t<m>olok* ***do***
3SG.NOM <AV>teach **1SG.OBL**
'He/She will be the one to teach me.'

There is a minor exception regarding the pronouns. As in many Austronesian languages, eight distinct person forms occur in Tboli: in addition to singular and plural forms for first, second, and third person, there are also first dual and first plural inclusive forms. Seven of these eight person forms have the special oblique form illustrated above, the exception being first plural inclusive *tekuy*, which is used in all grammatical functions (Forsberg 1992: 22; Porter 1977: 35–36). The following discussion will show that in Type II and Type III systems of DOM, it is rarely the case that all personal pronouns have the same argument realization options. Similar exceptions for individual pronoun forms are generally not found for Type I systems.

DOM in Tboli is asymmetric as personal names and pronouns occur in a basic unmarked form that is found in most grammatical contexts, the oblique usage illustrated above representing the marked case.

A symmetric example of Type II DOM is attested in Toratán (Ratahan). Here, personal names and pronouns generally occur in one of three case forms: nominative, genitive and locative (Himmelmann and Wolff 1999: 31). These cases are overtly marked by proclitics (*i=*, *ni=*, and *si=*, respectively), except for the genitive of the singular pronouns which makes use of special enclitics. In non-subject undergoer function, personal names and pronouns are marked with locative *si*, as illustrated in (22).

(22) Toratán (Ratahan), DOM in pronouns and personal nouns (Himmelmann and Wolff 1999: 23–24)

a. *Isé nto sumúq nanilow* ***si*** *tonaqas*
3SG LK.NR AV:enter AV.PST:look_for **LOC** shaman
'He would be the one to enter (Kinilow) and look for the shaman.'

b. *Ngapey* ***siyaq***
AV:wave **LOC.1SG**
'(He) beckoned me over.'

Common nouns always occur in their bare form without case markers when functioning as subjects or as non-subject undergoers regardless of their specificity.

5.3 Type III: DOM for pronouns only (Kimaragang, Indonesian)

DOM for pronouns only occurs in two different variants, one found primarily in Philippine-type languages, the other in non-Philippine-type languages. We begin with the former, using the Borneo language Kimaragang as our main example.

Basic clause structure in Kimaragang is similar to Tagalog and other Philippine-type languages: verbs are marked for symmetrical voice, and core arguments are overtly marked by one of three case proclitics (nominative, genitive or dative). The major difference from Tagalog is that all non-subject core arguments, except for recipients, receive genitive marking (Kroeger 2005: 406–407). Example (23a) illustrates genitive marking for an indefinite non-subject undergoer, (23b) for a definite non-subject undergoer, and (23c) for a personal name in non-subject undergoer function.

(23) Kimaragang, DOM for pronouns only

a. *Mangalapak okuh* ***do niyuw***
m-poN-lapak okuh **do niyuw**
AV-TR-split 1SG.NOM GEN coconut
'I will split a coconut/some coconuts.' (Kroeger 2005: 407)

b. *Ontok nopoh dit tanak dot [minonibas* ***dit tidi*** ***yoh]*** …
about only GEN child REL PST.AV.TR:slash **GEN** **mother** **3SG.GEN**
'As for the son who slashed (i.e. murdered) his mother, …' (Kroeger 2005: 413)

c. *Waro noh tulun sirih [dot sinumambat* ***di*** ***Majabou]***
exist FOC person ANAPH.LOC REL PST.AV:meet **GEN** **Majabou**
'There were people there who met Majabou …' (Kroeger 2005: 413)

Only pronouns in non-subject undergoer function require dative marking, as seen in (24).

(24) Kimaragang, dative form for pronouns in non-subject undergoer function

Kadung	*aa*	*kou*	*pendakod*	***dogon***,	*tibas-on*	*tekoo*
Kadung	aa	kou	Ø-po-indakod	**dogon**,	tibas-on	kuh-ikoo
if	NEG	2PL.NOM	AV-CAUS-climb	**1SG.DAT**	slash-PV	1SG.GEN-2PL.NOM

'If you don't let me climb up there, I'll slash you all to pieces!' (Kroeger 2005: 422)

Turning to non-Philippine-type languages showing Type III DOM, there are two crucial differences from the Kimaragang case just discussed. First, non-Philippine-type languages generally do not have overt case marking for common nouns in core argument functions (cf. Section 4.1), and the rare instances of overt case marking for personal names are restricted to the genitive. Second, DOM is not only restricted to personal pronouns in general, but rather more specifically to singular pronouns, and in some cases even to just the singular third person pronoun. Our main example is Indonesian, the standard variety of Malay spoken in Indonesia.

Indonesian generally does not show evidence for DOM. Pronouns, personal names, and common nouns in non-subject undergoer function usually occur in their base form, which is also used in practically all other grammatical functions such as subject, complement of preposition or possessor. In example (25), the first person singular pronoun *aku*[14] is seen in AV actor, AV undergoer and possessor functions, respectively.

(25) Indonesian

a. ***Aku*** *me-lihat* *guru=nya* *tadi.*
1SG AV-see teacher=3SG.GEN recently
'I saw the teacher earlier (today).'

b. *Guru=nya* *me-lihat* ***aku*** *tadi.*
teacher=3SG.GEN AV-see **1SG** recently
'The teacher saw me earlier today.'

c. *Ini* *pena* ***aku***.
PRX pen **1SG**
'This is my pen.' (Sneddon et al. 2010: 171)

However, singular pronouns (first, second, and third person) also have a short, clitic form which primarily occurs in possessor function, as in (26).

(26) Indonesian
*rumah=**ku**,* *mobil=**mu**,* *buku=**nya***
house=**1SG.GEN** car=**2SG.GEN** book=**3SG.GEN**
'my house, your car, her/his book' (Sneddon et al. 2010: 171)

14 There are in fact two forms of the first person singular pronoun, *aku* being the "more intimate form" when compared to the less intimate *saya* (Sneddon et al. 2010: 165).

This short form also occurs in non-subject undergoer function (and after a small number of prepositions). As a consequence, there are two alternative constructions when the undergoer in an AV construction is a singular pronoun, one with a full pronoun and one with a clitic pronoun, as seen in (27).

(27) Indonesian

a. *Narti menunggu **aku**.*
PN AV:wait **1SG**
'Narti is waiting for me.' (Sneddon et al. 2010: 171)

b. *Narti menunggu=**ku**.*
PN AV:wait=**1SG.GEN**
'Narti is waiting for me.' (Sneddon et al. 2010: 170)

There are thus two different marking options for pronominal non-subject undergoers, which formally constitutes an instance of DOM.

The functional difference between the two constructions, however, is not well understood. Sneddon et al. (2010) offer a few remarks on the use of the third person singular clitic *=nya*, which has a broad range of functions in Indonesian grammar, going well beyond the 'normal' scope of a possessive pronoun (e.g., use in nominalizations and for optional definiteness marking in NPs). Two points are of particular interest in the current context. First, unlike the third person singular free form *dia*, the clitic *=nya* may refer to inanimates, as shown in (28).

(28) Indonesian
*Surat itu, saya belum menerima=**nya**.*
letter DIST 1SG not.yet AV:receive=**3SG.GEN**
'Concerning that letter, I haven't received it yet.' (Sneddon et al. 2010: 290)

This usage shows that *=nya* appears to be more grammaticalized than its free counterpart *dia*. Second, example (28) also illustrates one typical usage context for the clitic, called "object topic-comment clause" by Sneddon et al. (2010: 290). Another very common usage is in object relative clauses (cf. also Arka 2021), as seen in (29).

(29) Indonesian
*sebuah lagu yang barangkali saudara akan menyukai=**nya***
one:CLF song REL apparently 2.HON will AV:like:APPL=**3SG.GEN**
'a song which perhaps you will like' (Sneddon et al. 2010: 298)

Given these observations, it seems likely that the difference between the two constructions may relate to the topicality and/or prominence of the undergoer. In the clitic construction, the antecedent for the pronoun is typically mentioned in the immediately preceding context. Its referent is thus typically highly activated and

discourse prominent. DOM here may thus be of the information structural type briefly mentioned in Section 2.2.

While details vary, similar alternations of full and clitic forms in non-subject undergoer function are attested across a range of non-Philippine-type languages. Totoli, a language spoken in Sulawesi, shows essentially the same distribution as Indonesian. In West Coast Bajau (northern Borneo), the alternation between free pronoun and genitive clitic for non-subject undergoers is only available for third person singular (obligatorily for 3rd person inanimate undergoers, optionally for animates). In Besemah (southern Sumatra), the alternation is also only found in the third person. In Karo Batak and Toba Batak (northern Sumatra), the genitive clitic is the only available option for third person singular non-subject undergoers. Strictly speaking, West Coast Bajau, Besemah and the two Batak languages thus constitute counterexamples to the specificity scale, as they show DOM (optionally or obligatorily) in third person (singular) pronouns, but not in first and second person.

5.4 Summary

Table 5 provides a comparison of the DOM assessments for WMP symmetrical voice languages in the literature (cf. Table 1 in Section 3) with our findings. The main result is clear: according to our assessments, DOM is found in more than half of the WMP languages listed in this table (13 out of 21), and hence more frequently than reported in the literature. As already noted in Section 3, our results probably still under-report the occurrence of DOM in WMP languages, because the available sources do not always provide sufficient evidence for deciding the issue (in addition to often being difficult to interpret).

The differences between our results and the ones reported in the literature depend of course in part on how DOM is defined, and which DOM-related phenomena are included in a particular database. For example, Type III DOM, which is limited to pronouns, is explicitly excluded by Sinnemäki (2014); hence the differing assessments for Indonesian and Riau Indonesian. Still, we would hold that the major factor underlying the differing assessments relates to the symmetrical voice systems characteristic for WMP languages, which makes it more difficult to identify DOM.

In this regard, it is important to note that all the phenomena that characterize typical DOM systems such as the Iranian ones presented in Section 2 also occur in WMP languages. Most importantly, the fact that the attestations of DOM in WMP languages almost perfectly adhere to the specificity scale provides strong support for the view that we are actually dealing with DOM. Furthermore, many instances of variation are similar to the ones documented for Iranian. For example, many central Philippine languages such as Tagalog and Hiligaynon provide clear evidence for

Table 5: WMP symmetrical voice languages mentioned in typological DOM surveys, classified according to the DOM types used for the current study.

Language	Iemmolo (2011)	Sinnemäki (2014) [non-pronominal asymmetric DOM only]	Witzlack-Makarevich et al. (2012)	This study
Hiligaynon	+DOM			+ Symmetric Type I
Malagasy	+DOM	+DOM		+ Asymmetric Type I
Tagalog	+DOM	−DOM		+ Symmetric Type I
Begak-Ida'an	+DOM	−DOM		+ Asymmetric Type I
Manadonese	+DOM	+DOM		+ Asymmetric Type I
Tboli		+DOM		+ Asymmetric Type II
Limos Kalinga		−DOM		−DOM
Kapampangan		−DOM		+ Asymmetric Type I
Tagabawa		−DOM		+ Asymmetric Type II
Indonesian		−DOM		+ Asymmetric Type III
Indonesian (Riau)		−DOM		+ Asymmetric Type III
Minangkabau		−DOM		−DOM
Sundanese		−DOM		−DOM
Batak (Toba)			−DOM	+ Asymmetric Type III
Balangao			−DOM	+ Symmetric Type II
Bontok			−DOM	−DOM
Gorontalo			−DOM	+ Asymmetric Type II
Ilocano			−DOM	−DOM
Javanese			−DOM	−DOM
Madurese			−DOM	−DOM
Mualang			−DOM	−DOM

DOM, while closely related Cebuano and Bikol do not (see also map in Appendix 1). Similar observations hold with regard to the types of systems (I, II or III) and their symmetry. Another similarity to the Iranian languages pertains to the fact that DOM forms of pronouns are more often fused when compared to DOM forms of common nouns and personal names. Likewise, some languages use the same DOM marker across different parts of speech; others have special forms for different parts of speech.

A more comprehensive table including all WMP symmetrical voice languages investigated for this study is given in Appendix 1. It supports our claim that DOM is robustly attested among WMP symmetrical voice languages. Philippine-type languages, especially the ones found in the central Philippines, tend to have Type I systems. A major exception is Philippine languages spoken in the northern half of Luzon which often lack DOM, contributing to their repeatedly noted special status among Philippine languages. Non-Philippine-type languages tend to have no DOM,

with the conspicuous exception of Malayic varieties, both standard national language varieties and Trade Malay varieties. Otherwise, Type II and Type III DOM systems are primarily found in the transitional area between Philippine-type and non-Philippine-type languages (Mindanao in the southern Philippines, northern Borneo, and northern Sulawesi; see the map in Appendix 1).

6 Discussion and conclusion

In Section 5 we have shown that, given the clarifications and definitions proposed in Section 4, DOM is robustly attested among WMP symmetrical voice languages, a finding not well reflected in typological databases for DOM. There are several factors contributing to the relative neglect and partial misrepresentation of DOM in WMP languages in such databases. The fact that the concept DOM is not yet a part of standard WMP grammaticography and the dearth of specialist DOM studies dealing with these languages is certainly of major relevance. Another major reason is that it is not a straightforward task to define DOM for symmetrical voice languages, as is done here in Section 4.

These difficulties, to our minds, have implications for large-scale typological database projects striving to gather data points for a particular phenomenon across the world's languages. The applicability of data gathering strategies varies significantly in line with the phenomenon investigated. Some phenomena allow for simple coding routines that can be partially automatized, provided enough electronic grammars are available. Other phenomena require specialist analyses as a prerequisite for deciding on the values to be entered into the database for a particular language. The latter are probably necessary for those phenomena that are deeply embedded in the morphosyntactic system of a language. Grammatical relations and associated phenomena such as DOM are primary examples. Other phenomena, such as the basic system of person distinctions in pronouns, tend to be less deeply embedded in the overall structure of a given language, and hence more amenable to a simple coding routine.

In Section 1, we distinguished a universalist view on DOM from a contact view on DOM, the former holding that DOM phenomena are robustly attested across the world's languages, rendering a genealogical or contact-based explanation of their distribution not very likely. The findings in Section 5 support the universalist view in two regards. First, they provide strong support for the specificity scale which predicts DOM to always include pronouns and personal names when it is also attested for common nouns. The specificity scale is also found by Bickel et al. (2015: 33, 38) to be the most likely candidate for a universal scale effect on case marking.

In this regard, the current study also confirms another important finding by Bickel et al. (2015: 9, 13–14): personal pronouns do not behave uniformly regarding DOM. While it is true that if a language has DOM, DOM is (almost?) always found for pronouns, it is not the case that all pronouns in a given system show DOM. In WMP symmetrical voice languages, plural personal pronouns sometimes do not participate in the DOM system, which is true in particular for Type II and Type III systems (Sections 5.2 and 5.3). Furthermore, the behavior of third (singular) pronouns is variable: sometimes they align with first and second singular pronouns and show DOM; sometimes they are the only category in a language to which DOM applies (cf. West Coast Bajau, Besemah and Batak in Section 5.3). In short, while the overall structure of the specificity scale appears to be robustly attested across the languages of the world, the details for personal pronouns are variable and cannot be captured by a single scale of the type: first/second singular > third singular > non-singular pronouns.

The second regard in which the current findings support the universalist position pertains to the fact that DOM is well attested in an area that has been represented as being "DOM-free". Bickel et al. (2015: 28) argue that there are two "frequency peaks" among the languages of the world in terms of DOM. One – called "Eurasia" – is centered on the Indo-Iranian speaking territories, but it reaches to Europe (Romance) as well as Inner Asia (Uralic, Turkic, Mongolic, Sino-Tibetan), and it also includes the Indian subcontinent (Dravidian). The other "frequency peak" of DOM is located (loc. cit.) "in the New-Guinea/Australia – or 'Sahul' – macroarea (centered on Pama-Nyungan languages but extending to Tangkic and Southern New Guinea)." In this view, there is a major gap in between these two macro-areas constituted by DOM-lacking languages in continental Southeast Asia, Indonesia, and the Philippines. This gap, however, is considerably lessened if Indonesia and the Philippines (and Madagascar) have their DOM instances as well, as it basically means that DOM is robustly attested in an area stretching across three continents, too large an area to be likely to warrant an explanation in terms of contact.

Acknowledgment: We thank Sander Adelaar, Russel Barlow, Paul Kroeger, Jozina Vander Klok and Taras Zakharko for help with the data, and Maria Bardají for creating the map. We also thank three anonymous reviewers and the editorial board for their questions and suggestions which have helped to sharpen our argument and exposition. The research for this article was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252 (CRC *Prominence in Language*).
Author contributions: All authors participated in developing hypotheses and the overall argument. SB compiled the data base and wrote the first draft of all sections except Section 4. SR drafted Section 4 and contributed to the revision of all other

sections. NH prepared the second and third drafts of the complete manuscript. NH and SR expanded, checked and revised data coding.

Data availability statement: Data supporting the results reported in this article (language names, subgroup affiliation, sources consulted, and type of DOM, if applicable) can be found in Appendix 1.

Abbreviations

1	first person
2	second person
3	third person
ABL	ablative
ACC	accusative
APPL	applicative
ASP	aspect
AV	actor voice
CAUS	causative
CLASS	classifier
COMP	complementizer
CONJ	conjunctive
CONT	continuative (aspect)
CPL	completive
CV	conveyance voice
DAT	dative
DEM	demonstrative
DET	determiner
DIR	directional
DIST	distal
DU	dual
ERG	ergative
EXCL	exclusive
EXIST	existential
FOC	focus
GEN	genitive
HON	honorific
IND	indicative
LK	linker
LOC	locative
LV	locative voice
MED	medial
NEG	negation, negative
NEUT	neuter
NOM	nominative
NR	nominalizer

NSPEC	not specific
OBL	oblique
PL	plural
PN	personal name
POT	potential(is)
PRET	preterite
PRS	present
PRX	proximal
PST	past
PTCP	participle
PV	patient voice
REL	relative
RLS	realis
SG	singular
SPEC	specific
ST	stative
TR	transitive
UNF	unfocussed
UV	undergoer voice

References

Abaev, Vasilij Ivanovič. 1964. *A grammatical sketch of Ossetic* [= Grammatičeskij očerk osetinskogo jazyka, 1959, Ed. by Herbert H. Paper. Transl. by Steven P. Hill]. Bloomington: Indiana University.

Adelaar, Alexander. 2005a. Structural diversity in the Malayic subgroup. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 202–226. London and New York: Routledge.

Adelaar, Alexander. 2005b. The Austronesian languages of Asia and Madagascar: A historical perspective. In Alexander Adelaar & Nikolaus Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 1–41. London and New York: Routledge.

Adelaar, Alexander & Nikolaus P. Himmelmann (eds.). 2005. *The Austronesian languages of Asia and Madagascar*. London and New York: Routledge.

Adelaar, Alexander & Ritsuko Kikusawa. 2014. Malagasy personal pronouns: A lexical history. *Oceanic Linguistics* 53(2). 480–516.

Aissen, Judith. 2003. Differential object marking: Iconicity versus economy. *Natural Language and Linguistic Theory* 21. 435–483.

Arka, I Wayan. 2021. Pivot and puzzling relativization in Indonesian. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar*, 181–202. Oxford: OUP.

Bárány, András & Laura Kalin. 2020. Introduction. In András Bárány & Laura Kalin (eds.), *Case, agreement, and their interactions. New perspectives on differential argument marking*, 1–25. Berlin, Boston: de Gruyter.

Benton, Richard A. 1971. *Pangasinan reference grammar*. Honolulu: University of Hawaii Press.

Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko. 2015. Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky,

Andrej L. Malchukov & Marc D. Richards (eds.), *Scales and hierarchies: A cross-disciplinary perspective*, 7–43 (Trends in Linguistics. Studies and Monographs 277). Berlin: de Gruyter Mouton.

Blust, Robert. 2002. Notes on the history of 'focus' in Austronesian languages. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 63–78. Canberra: Pacific Linguistics.

Blust, Robert. 2015. The case-markers of Proto-Austronesian. *Oceanic Linguistics* 54(2). 436–491.

Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr. Available at: https://zenodo.org/record/4697660.

Bossong, Georg. 2021. DOM and linguistic typology. A personal view. In Johannes Kabatek, Philipp Obrist & Albert Wall (eds.), *Differential object marking in Romance: The third wave*, 21–46. Berlin, Boston: de Gruyter.

Chappell, Hilary & Jean-Christophe Verstraete. 2019. Optional and alternating case marking: Typology and diachrony. *Language and Linguistics Compass* 13(3). https://doi.org/10.1111/lnc3.12311.

Chen, Victoria & Bradley McDonnell. 2019. Western Austronesian voice. *Annual Review of Linguistics* 5. 173–195.

Christensen, Arthur. 1915. *La dialecte du Sämnān. Essai d'une grammaire sämnānie avec un vocabulaire et quelques textes, suivi d'une notice sur les patois de Sängsar et de Lāsgird*. [The dialect of Sämnān, essay about a Sämnāni grammar with a vocabulary and some texts, followed by a note on the dialects of Sängsar and Lāsgird]. København: Høst.

Christensen, Arthur. 1935. *Contributions à la dialectologie iranienne II. Dialectes de la région de Sèmnān: Sourchéī, Lāsguerdī, Sängesärī et Chämerzådī* [Contributions to an Iranian dialectology II. Dialects of the region of Sèmnān: Sourchéī, Lāsgerdī, Sängesärī and Chämerzådī]. København: Levin & Munksgaard.

Compensis, Paul. 2022. *Differential object indexing in Bulgarian – The role of discourse prominence and predictability*. Cologne: University of Cologne PhD dissertation. Available at: https://kups.ub.uni-koeln.de/62893/.

Comrie, Bernard. 1979. Definite and animate direct objects: A natural class. *Linguistica Silesiana* 3. 13–21.

Croft, William. 1988. Agreement versus case marking and direct objects. In Michael Barlow & Charles A. Ferguson (eds.), *Agreement in natural language. Approaches, theories, descriptions*, 159–179. Stanford: Center for the Study of Language and Information Publications.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

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

de Hoop, Helen & Andrej L. Malchukov. 2008. Case-marking strategies. *Linguistic Inquiry* 39. 565–587.

de Swart, Peter. 2007. *Cross-linguistic variation in object marking*. Nijmegen: Radboud University Nijmegen dissertation.

Foley, William A. 1998. Symmetrical voice systems and precategoriality in Philippine Languages. Paper presented at the 3rd LFG conference, University of Queensland, 30 June – 3 July. Available at: https://zenodo.org/records/5336774.

Foley, William A. 2008. The place of Philippine languages in a typology of voice systems. In Peter K. Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*, 22–44. Stanford: Center for the Study of Language and Information Publications.

Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Forsberg, Vivian M. 1992. A pedagogical grammar of Tboli. *Studies in Philippine Linguistics* 9(1). 1–110.

Grjunberg, Aleksandr L. 1972. *Jazyki vostočnogo Gindukuša. Mundžanskij jazyk. Teksty, slovar', grammatičeskij očerk* [Languages of the eastern Hindukush. The Munji language. Texts, lexicon, grammatical main features]. Leningrad: Nauka.

Haig, Geoffrey. 2018. The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics* 56(4). 781–818.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London and New York: Routledge.

Himmelmann, Nikolaus P. 2008. Lexical categories and voice in Tagalog. In Peter K. Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*, 247–293. Stanford: Center for the Study of Language and Information Publications.

Himmelmann, Nikolaus P. 2021. Notes on 'noun phrase structure' in Tagalog. In Jens Fleischhauer, Anja Latrouite & Rainer Osswald (eds.), *Explorations of the syntax-semantics interface*, 319–342. Berlin, Boston: Düsseldorf University Press.

Himmelmann, Nikolaus P. & John U. Wolff. 1999. *Toratán (Ratahan)*. Munich, Newcastle: LINCOM Europa.

Hopperdietzel, Jens. 2020. Pseudo noun incorporation and differential object marking: Object licensing in Daakaka. In Ileana Paul (ed.), *Proceedings of the twenty-sixth meeting of the Austronesian Formal Linguistics Association (AFLA)*, 106–123. Ontario: University of Western Ontario.

Howe, Penelope. 2022. Semantics and pragmatics of voice in Central Malagasy oral narratives. *Oceanic Linguistics* 61. 68–117.

Iemmolo, Giorgio. 2011. *Towards a typological study of differential object marking and differential object indexation*. Pavia: University of Pavia dissertation. Available at: https://zenodo.org/record/4388380/files/dom_iemmolo_ms.pdf.

Iemmolo, Giorgio. 2013. Symmetric and asymmetric alternations in direct object encoding. *STUF – Language Typology and Universals* 66(4). 378–403.

Josephs, Lewis S. 1984. The impact of borrowing on Palauan. In Byron W. Bender (ed.), *Studies in Micronesian linguistics*, 81–123. Canberra: ANU Asia-Pacific Linguistics. Available at: http://hdl.handle.net/1885/253384.

Kagan, Olga. 2020. Differential object marking. In Olga Kagan (ed.), *The semantics of case* (Key Topics in Semantics and Pragmatics), 147–188. Cambridge: Cambridge University Press.

Kalin, Laura. 2018. Licensing and differential object marking: The view from Neo-Aramaic. *Syntax* 21(2). 112–159.

Kaufman, Daniel. 2009. Austronesian nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35. 1–49.

Keenan, Edward L. 2008. The definiteness of subjects and objects in Malagasy. In Greville Corbett & Michael Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie* (Typological Studies in Language, Volume 81), 241–261. Amsterdam, Philadelphia: John Benjamins.

Kieffer, Charles M. 2009. Parachi. In Gernot Windfuhr (ed.), *The Iranian languages*, 693–720. London, New York: Routledge.

Kroeger, Paul. 1993a. Another look at subjecthood in Tagalog. *Philippine Journal of Linguistics* 24(2). 1–15.

Kroeger, Paul. 1993b. *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: CSLI Publications.

Kroeger, Paul. 2005. Kimaragang. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 397–428. London and New York: Routledge.

Kroeger, Paul & Sonja Riesberg. 2024. Voice and transitivity. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of South East Asia*, 793–821. Oxford: Oxford University Press.

Latrouite, Anja. 2011. Differential object marking in Tagalog. In Lauren Eby Clemens, Gregory Scontras & Maria Polinsky (eds.), *Proceedings of the eighteenth meeting of the Austronesian Formal Linguistics Association (AFLA)*, 94–109. Cambridge, MA: Harvard University.

Lee, Nala H. 2022. *A grammar of modern Baba Malay* (Mouton Grammar Library, Vol. 90). Berlin, Boston: de Gruyter Mouton.

Mardale, Alexandru & Petros Karatsareas. 2020. Differential object marking and language contact: An introduction to this special issue. *Journal of Language Contact* 13(1). 1–16.

McDonnell, Bradley & Victoria Chen. 2022. The evolution of syntax in western Austronesian. In Chris Shei & Saihong Li (eds.), *The Routledge handbook of Asian linguistics*, 11–32. London; New York: Routledge.

McFarland, Curtis D. 1978. Definite objects and subject selection in Philippine languages. In Casilda Edrial-Luzares & Austin Hale (eds.), *Studies in Philippine linguistics*, vol. 2, 139–182. Manila: Linguistic Society of the Philippines.

Mithun, Marianne. 1994. The implications of ergativity for a Philippine voice system. In Barbara A. Fox & Paul J. Hopper (eds.), *Voice: Form and function*, 247–277. Amsterdam, Philadelphia: John Benjamins.

Moravcik, Edith A. 1978. On the case marking of objects. In Joseph Greenberg (ed.), *Universals of human language, Volume 4: Syntax*, 249–290. Stanford: Stanford University Press.

Morgenstierne, Georg. 1929. *Indo-Iranian frontier languages I. Parachi and Ormuri*. Oslo: Aschehoug.

Næss, Åshild. 2004. What markedness marks: The markedness problem with direct objects. *Lingua* 114(9–10). 1186–1212.

Pachalina, Tat'jana Nikolaevna. 1969. *Pamirskie jazyki* [Pamirian languages]. Moskva: Nauka.

Pearson, Matthew. 2001. *The clause structure of Malagasy: A minimalist approach*. Los Angeles: University of California dissertation.

Porter, Doris. 1977. *A Tboli grammar* (Linguistic Society of the Philippines Special Monograph Issue 7). Manila: Linguistic Society of the Philippines.

Richards, Charles Monroe. 1971. *A case grammar of Pampangan*. Los Angeles: University of California dissertation.

Riesberg, Sonja. 2014. *Symmetrical voice and linking in Western Austronesian languages* (Pacific Linguistics, Vol. 646). Berlin, Boston: de Gruyter Mouton.

Riesberg, Sonja, Kurt Malcher & Nikolaus P. Himmelmann. 2019. How universal is agent-first? Evidence from symmetrical voice languages. *Language* 95(3). 523–561.

Rodríguez-Ordóñez, Itxaso. 2017. Reexamining differential object marking as a linguistic contact-phenomenon. *Journal of Language Contact* 10(2). 318–352.

Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above? In Charles Li (ed.), *Subject and topic*, 493–518. New York: Academic Press. Available at: https://www.ling.uni-potsdam.de/~thiersch/Malagasy/schachter_4934_fixed.pdf.

Schapper, Antoinette. 2015. Wallacea, a linguistic area. *Archipel. Études interdisciplinaires sur le monde insulindien* 90. 99–151.

Schmidtke-Bode, Karsten & Natalia Levshina. 2018. Reassessing scale effects on differential case marking: Methodological, conceptual and theoretical issues in the quest for a universal. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 509–537. Berlin: Language Science Press.

Shibatani, Masayoshi. 1988. Introduction. In Masayoshi Shibatani (ed.), *Passive and voice*, 1–8. Amsterdam, Philadelphia: John Benjamins.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert Malcolm Ward Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Sinnemäki, Kaius. 2014. A typological perspective on differential object marking. *Linguistics* 52(2). 281–313.

Smith, Alexander D. 2017. The Western Malayo-Polynesian problem. *Oceanic Linguistics* 56(2). 435–490.
Sneddon, James Neil, Alexander Adelaar, Dwi Djenar & Michael Ewing. 2010. *Indonesian: A comprehensive grammar*, 2nd edn. London: Routledge.
Sokolova, Valentina S. 1973. *Genetičeskie otnošenija mundžanskogo jazyka i šugnano-jazguljamskoj jazykovoj gruppy* [The genetic relationship between the Munji language and the Shughni-Yazghulami language group]. Leningrad: Nauka.
Spitz, Walter L. 2001. *Hiligaynon/Ilonggo* (Languages of the World/Materials 209). München: Lincom Europa.
Spitz, Walter L. 2002. Voice and role in two Philippine languages. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of Western Austronesian voice systems* (Pacific Linguistics 518), 379–404. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
Witzlack-Makarevich, Alena, Lennart Bierkandt, Taras Zakharko & Balthasar Bickel. 2012. Case alignment. https://www.autotyp.uzh.ch/available.html (accessed 27 May 2024).
Witzlack-Makarevich, Alena & Ilja A. Seržant. 2018. Differential argument marking: Patterns of variation. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 1–40. Berlin: Language Science Press.
Zobel, Erik. forthcoming. Historical linguistics of the Central Malayo-Polynesian languages. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia*. Oxford: OUP.
Zribi-Hertz, Anne & 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 & Seppo Kittilä. 2019. *Grammatical voice*. Cambridge: Cambridge University Press.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/lingty-2024-0002).