



UvA-DARE (Digital Academic Repository)

The expression of modifiers and arguments in the noun phrase and beyond

A typological study

van Rijn, M.A.

Publication date

2017

Document Version

Other version

License

Other

[Link to publication](#)

Citation for published version (APA):

van Rijn, M. A. (2017). *The expression of modifiers and arguments in the noun phrase and beyond: A typological study*. LOT.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

4

Alignment beyond the clause: the morphosyntactic expression of phrasal and clausal dependency relations from a typological perspective¹

Abstract

This paper investigates alignment, not only as a clausal phenomenon, but as applying to both clauses and phrases. Specifically, I investigate the degree of identical morphosyntactic coding of possessive NPs, adpositional phrases and arguments in verbal main clauses, in a worldwide sample of 39 languages. It is shown that alignment patterns are sensitive to a semantic distinction between modifier dependents and argument dependents on the one hand, and to the phrasal versus clausal nature of the dependency relation on the other. This finding demonstrates the typological relevance of alignment as a phenomenon that transcends the domain of the clause.

4.1 Introduction

Ever since Nichols' (1986) seminal paper, it has been known that the syntactic relationship between two units in a phrase or clause can be analyzed in terms of head and dependent. The head governs, or subcategorizes for, the dependent and determines the syntactic category and distribution of the constituent as a whole. This highly influential distinction, developed within structuralist syntax and, more specifically, within the framework of dependency grammar (e.g. Tesnière 1959; Mel'čuk 1988), applies to a range of basic phrase and clause types, of which possessive noun phrases, adpositional phrases and the relationship between a verb and its core argument(s) in main clauses are relevant to this paper.

This paper starts from the observation that dependency relations may be of two semantic types. On the one hand, there are heads that inherently require reference to a dependent, which may therefore be considered the head's *argument*. Typical examples of such inherently relational heads are adpositions

¹ This chapter has been submitted for publication.

and verbs, but also nouns denoting, for example, kinship terms and body parts. On the other hand, there are heads that take their dependent as a *modifier*; such dependents are not inherently presupposed by the head, but merely provide additional information about it. Examples of such dependency relations are those between a noun and an attributive adjective, a verb and a (manner/degree) adverb, as well as those between prototypical non-relational nouns, such as those denoting concrete inanimate objects, and their possessor. Importantly, the distinction is relevant to the range of phrasal and clausal dependency relations investigated by Nichols, as it divides these relations into those containing arguments and those containing modifiers.

Languages show extensive variation in the degree to which they have dedicated means of coding – in the form of indexing or flagging² – for different dependency relations. As demonstrated by Siewierska (1998), formal correspondences between index-sets are cross-linguistically common: 52% (82 out of 157) of the languages in her sample use (fully or partially) identical person forms for possessors on nouns and for arguments on verbs. Similarly, languages commonly exhibit identities in inflectional case forms – a phenomenon known as ‘case syncretism’ (Baerman et al. 2005: 38–57; Baerman 2008). This obtains not only within clauses, i.e. syncretism of core grammatical cases (nominative with accusative, and absolutive with ergative), but also across phrases and clauses. An example is the use of a single genitive/ergative case form for adnominal possessors and A arguments, as in West Greenlandic and other Eskimo languages.³ Gil (2013) furthermore demonstrates that it is common for languages to use the same means of coding for different noun modifiers, more specifically possessors, attributive adjectives and relative clauses.

This paper investigates to what extent languages use identical means of coding for a range of phrasal and clausal dependency relations. Specifically, the aim is to show to what extent patterns of identical coding are sensitive to the semantic distinction between modifier dependents and argument dependents on the one hand, and the phrasal versus clausal nature of the dependency relation

² I follow Haspelmath (2005, 2013a) in using the term *index(ing)* for affixal person forms (irrespective of whether they are referential markers or agreement markers, cf. Chapters 2 and 3 of this thesis) and the term *flagging* for case markers, adpositions, attributive particles and other invariant markers of the dependency relation. Purely morphological marking, such as tonal marking, is also subsumed under the notion of flagging.

³ I use the following labels for verbal argument roles: S_A is the agent-like argument of a one-participant predicate, S_P is the patient-like argument of a one-participant predicate, A is the more agent-like argument of a two-participant predicate, and P is the more patient-like argument of a two-participant predicate.

on the other. To this end, I examine the morphosyntactic coding of three classes of dependency relations in a worldwide sample of 39 languages: modifiers within phrases (possession of non-relational nouns), arguments within phrases (possession of relational nouns, adpositional phrases) and arguments within main clauses (the core arguments of one- and two-participant verbs).⁴ I hypothesize that one logically possible pattern of coding is cross-linguistically rare or non-existent: one in which modifiers in phrases receive the same morphosyntactic treatment as arguments in clauses, while arguments in phrases are treated differently.

A problem in comparing languages in terms of the coding of dependency relations is that a single dependency relation may be expressed by multiple means of coding. This phenomenon has been discussed most extensively in typological work on clausal alignment, where it has received a number of different labels including ‘split’ coding (Silverstein 1976), ‘differential subject/object marking’ (Bossong, 1985, 1998; Comrie 1989) and ‘differential argument marking’ (Baerman 2008). However, it also plays a crucial – yet understudied – role in phrasal dependency relations, including those investigated in this study. A major source of differential coding patterns are referential properties of dependents, including lexical properties, such as person, humanness, animacy and (pro)nominality, and discourse-based properties, such as definiteness and specificity (Bickel 2010). In English, for instance, the choice of genitive construction is in part conditioned by the referential type of possessor: the ‘Saxon’ genitive (*John’s house*) is preferred for human nouns and nouns denoting higher animals, while the ‘Norman’ genitive (*house of John*) is preferred for collective nouns denoting groups of people (see Keizer 2007 for a comprehensive overview of the relevant factors). In this paper, I take into account the full range of variation in the coding of phrasal and clausal dependency relations conditioned by referential factors in individual languages.⁵ Moreover, I show how languages can be compared in the face of such variation, following much recent work in clausal alignment typology (Bickel 2010; Bickel et al. 2013; Bickel et al. 2015; Witzlack-Makarevich 2011).

⁴ Considerations concerning the choice of these specific constituent types are provided in Section 4.4.2.

⁵ Other factors that may affect the coding of dependency relations are properties of the head, such as the lexical type of noun or verb, and properties of the whole clause, such as tense/aspect/mood. The ways in which these factors are dealt with in the present study are discussed in Section 4.4.2.

The organization of the paper is as follows: Section 4.2 discusses the modifier/argument opposition, as pertaining to dependency relations of different kinds, in more detail. In Section 4.3, I formulate the hypothesis that follows from this theoretical opposition, tested in this study. Section 4.4 provides the necessary methodological background: Section 4.4.1 presents the language sample, and Section 4.4.2 discusses the types of dependency relations investigated. Section 4.4.3 illustrates how referential factors may impact morphosyntactic coding in clauses and, specifically, phrases, and shows how this variation is dealt with in the study. Section 4.4.4 provides further methodological preliminaries, and, finally, Section 4.4.5 explains how language-specific coding patterns relate to the hypothesis formulated in Section 4.3. The results of the study are presented in Section 4.5. Section 4.6 in turn discusses the central role of adpositional phrases in the overall patterns. Finally, Section 4.7 summarizes the main findings and provides directions for further research.

4.2 The modifier/argument distinction

In this paper, I start from the observation that dependency relations, as they are investigated by Nichols, may be of two semantic types: those between a head and its argument, and those between a head and its modifier. Heads that take their dependent as an argument are inherently relational, i.e. they encode concepts that by their very nature set up a relationship with a dependent. Well-known examples are adpositions and verbs. With a verb, this property is generally known as ‘valence’, i.e. “its inherent relationality that allows it to govern a particular number of arguments of a particular type” (Haspelmath & Müller-Bardey 2004: 1130). However, as pointed out by a number of scholars (Seiler 1983a; Lehmann 1985; Dahl & Koptjevskaja-Tamm 1998; Partee 1997; Partee & Borschev 2003), the same property also applies to nouns: relational, or bivalent, nouns are intrinsically linked to another item. Clear examples of such nouns are kinship terms and body part terms: while a noun such as ‘mother’ inherently denotes a relationship between two individuals, a noun such as ‘arm’ inherently denotes a part of an (animate) whole (Barker 1995; Taylor 1996).⁶ Both the presence and the type of dependent are determined by the lexical

⁶ These items can of course also be used in a non-relational sense, e.g. *John’s leg* can refer to a chicken leg that John is eating, but such interpretations are pragmatically marked and occur only in certain special contexts (cf. Lichtenberk et al. 2011: 671).

specification of the relational head noun, in a fashion reminiscent of verbal predicates and adpositions.⁷

By contrast, there are dependents that are not inherently presupposed by their head. They merely provide a further semantic characterization of the referent or state-of-affairs expressed by the head, thereby functioning as a modifier. Well-known examples of noun modifiers are attributive adjectives, relative clauses, and possessors of non-relational nouns. Typical non-relational nouns, such as ‘pot’ or ‘pen’, denote concrete (inanimate), countable objects. Such nouns lack an inherently relational meaning; as a result, the dependency relation may receive a range of different interpretations, determined by the context of use. A phrase such as *Mary’s pot*, for instance, can refer to a pot she made, a pot she bought, one that she is using, etc. (Seiler 1983a: 40–41; Koptjevskaja-Tamm 2004; Lichtenberk et al. 2011).

Interestingly, languages differ extensively in the degree to which the semantic type of dependency relation is reflected morphosyntactically, either in terms of indexing or in terms of flagging. Some languages make a clear distinction between modifiers on the one hand and arguments on the other. In the language isolate Burushaski (Pakistan), for instance, a single set of person prefixes is used on relational items, be they nominal (1a), verbal (1b), or adpositional (1c):

- (1) a. *a-yáʔis*
 1SG-head
 ‘my head’
- b. *a-pʰús-u*
 1SG-tie.up-3SG
 ‘she tied me up’
- c. *a-pʌčǐ*
 1SG-with
 ‘with me’ (Berger 1998: 91; Berger 1974: 49; Lorimer 1935: 96)

⁷ Other nouns that may fit this semantic profile are spatial nouns (e.g. ‘top’ or ‘bottom’), or those denoting parts of inanimate wholes (e.g. ‘branch’ or ‘handle’) and physical or mental states (e.g. ‘strength’ and ‘fear’). As is well known, the valency of individual lexical items, be they nouns or verbs, varies extensively cross-linguistically. The way in which the distinction between modifiers and arguments is made in the present study is discussed in Section 4.4.2.

Modifiers, including possessors of non-relational nouns, remain unindexed, as shown in (2):

- (2) *ǰã hághoor*
 1SG.GEN horse
 ‘my horse’ (Biddulph 1884: 22)

A formal opposition between possessive arguments, as in (1a), and possessive modifiers, as in (2), is cross-linguistically common, and generally known as a distinction between *alienable* and *inalienable possession*.

Means of flagging may also be sensitive to the modifier/argument opposition. This can be illustrated for Katukina-Kanamari (Katukinan, Brazil), in which a single genitive/ergative case suffix is used for arguments of nouns (3a), verbs (3b) and adpositions (3c):

- (3) a. *Pityira-na=tyo Tikon.*
 Pityira-GEN=daughter Tikon
 ‘Tikon is Pityira’s daughter.’
 b. *Kopa-na=ti: pi:da.*
 Kopa-ERG=killed jaguar
 ‘Kopa killed a jaguar.’
 c. *Pioru-na=katu Tirin.*
 Pioru-OPOSTP=com Tirin
 ‘Tirin is with Pioru.’ (Dos Anjos 2011: 224)

Possessive modifiers also take this case suffix, but additionally require the possessive marker *wa* to occur in between the modifier and the possessed noun, as demonstrated in (4):

- (4) *Kontan-na=wa poako itowun*
 Kontan-GEN=POSS paddle this
 ‘This is Kontan’s paddle.’ (Queixalós forthcoming: 6)

Additionally, modifiers and arguments may be distinguished by means of the presence vs. absence of flagging; in Bambara (Mande, Mali), for instance, arguments of nouns (5a), verbs (5b) and adpositions (5c) remain unmarked:

- (5) a. *mùso fâ*
 woman father
 ‘the woman’s father’
 b. *dònsow sé-ra*
 hunters arrive-PRF
 ‘the hunters arrived’
 c. *du kònò*
 compound inside
 ‘inside the compound’ (Hewson 2014: 6, 5, 63)

Possessive modifiers, however, require a possessive marker *ká*, as shown in (6):

- (6) *mùso ká líburu*
 woman POSS book
 ‘the woman’s book’ (Hewson 2014: 6)

Conversely, the distinction between modifiers and arguments may be formally neutralized to varying degrees. In Dutch (Indo-European, Netherlands), for instance, the preposition *van* is used not only for possessive modifiers (7a), but also for possessive arguments (7b) and for arguments of (a set of) prepositions (7c):

- (7) a. *het boek van Jan*
 DEF.N book of Jan
 ‘Jan’s book’
 b. *de vader van de jongen*
 DEF.C father of DEF.C boy
 ‘the father of the boy’
 c. *links van het huis*
 left of DEF.N house
 ‘left of the house’

A more extensive degree of formal neutralization is found in Abkhaz (Northwest Caucasian, Georgia), where a single set of prefixes is used on non-relational nouns (8a), relational nouns (8b), adpositions (8c) and for (1st and 2nd person) S arguments (8d), A arguments (8e), and P arguments (8f):

- (8) a. $s_9\text{-}\check{s}^{\circ}q^{\circ}\check{s}$
1SG-book
'my book' (Chirikba 2003: 57)
- b. $s\text{-}an$
1SG-mother
'my mother' (Hewitt 2008: 76)
- c. $s\text{-}q^{\circ}nt^{\circ}$
1SG-from
'from me' (Hewitt 1979: 103)
- d. $s\text{-}ca\text{-}wá\text{-}jt'$
1SG-go-PRS-FIN
'I am going.' (Chirikba 2003: 39)
- e. $s\text{-}j\check{\theta}\text{-}s(\check{\theta})\text{-}\emptyset\text{-}(j)t'$
1SG-3SG-hit-AOR-FIN
'I hit him.' (Chirikba 2003: 40)
- f. $d\check{\alpha}\text{-}z\text{-}s\text{-}'m\check{\alpha}\text{-}s\check{\alpha}\text{-}\emptyset\text{-}jt'$
3SG-1SG-POT-not-hit-PST-FIN
'S/he couldn't hit me.' (Hewitt 2008: 81)

The degree to which languages formally neutralize the modifier/argument distinction across phrases and clauses, as demonstrated in (1) to (8), is the focus of the present study. I henceforth refer to such patterns of formal neutralization as *alignment patterns*, following the traditional use of the term for the formal neutralization of (a subset of) core arguments in clauses.

Modifiers and arguments are found in both phrasal and clausal dependency relations, which yields four logically possible domains of morphosyntactic coding, outlined in Table 1 below:

Table 1: Four types of dependency relations

		Semantic relationship	
		Head-modifier	Head-argument
Syntactic category	Phrase	Modifier in phrase	Argument in phrase
	Clause	Modifier in clause	Argument in clause

One of the four types of dependency relations – modifiers in clauses – is excluded from this study, for three reasons. First, this class is internally quite diverse, as it includes both adjunct phrases and (manner/degree) adverbs. Likewise, the cross-linguistic coding of the two types of modifiers varies

significantly: while adjunct phrases typically receive different means of flagging, depending on their semantic role such as beneficiary, comitative or location, adverbs rarely take overt means of coding. Second, it proves to be very difficult to distinguish modifiers and arguments in clauses in individual languages (see e.g. Creissels 2014; Forker 2014; Haspelmath 2014; Haspelmath & Hartmann 2015); different tests of argumenthood have been proposed in the literature, but each is known to suffer from a range of limitations (cf. Comrie 1993). Finally, the exclusion of clause modifiers has a theoretical motivation: adjunct phrases modify a non-lexical head, i.e. the entire clause, which consist minimally of the verb and one or more core arguments, while the modifiers (and arguments) investigated in this study are dependents of a lexical head. This also applies to manner/degree adverbs, which may modify at the level of the clause (*John angrily left the room*) aside from modifying at the level of the individual predicate (*John walked slowly*).

This study thus limits itself to three classes of dependency relations: modifiers in phrases, arguments in phrases and arguments in clauses. The individual dependency relations in each class studied in the paper are discussed in Section 4.4.2. In the next section, I present the hypothesis tested in this study.

4.3 Hypothesis

This paper investigates to what extent languages have identical means of coding for different phrasal and clausal dependency relations; specifically, I expect patterns of coding to be constrained by the modifier/argument distinction on the one hand, and by the phrasal vs. clausal nature of the dependency relation on the other. This prediction yields a classification of five logically possible types of alignment patterns that may be found in individual languages, presented in Table 2 below. In this table, capitals indicate neutralization by means of a single coding strategy, be it indexing, flagging, or the absence of both, i.e. zero-coding.

Table 2: Typology of alignment types

	1	2	3	4	5
Modifier in phrase	A	A	A	A	A
Argument in phrase		B		B	B
Argument in clause	B			C	A

In alignment patterns of the first type, modifiers in phrases receive the same coding strategy as arguments in phrases, while arguments in clauses are

coded differently. In the second pattern, arguments in phrases receive the same coding strategy as arguments in clauses, while modifiers in phrases are coded differently. The third pattern illustrates the formal neutralization of dependents in each of the three categories, while the fourth pattern illustrates the inverse: the complete absence of formal neutralizations across the three types of dependency relations. The fifth logically possible pattern is one in which modifiers in phrases are coded like arguments in clauses, while arguments in phrases receive a different coding strategy.⁸ Following observations by Hengeveld & Mackenzie (2008: 383–387) and an earlier pilot study (Van Rijn 2011), I predict patterns 1 to 4 to be cross-linguistically common, while pattern 5 is expected to be cross-linguistically rare or non-existent. This prediction yields the following hypothesis, tested in this study:

- (i) If modifiers in phrases are coded in the same way as arguments in clauses, arguments in phrases receive the same formal treatment.

Put differently, coding strategies are expected to cover continuous segments of the following scale:

- (9) modifier in phrase – argument in phrase – argument in (main) clause

I expect languages to use a single coding strategy for modifiers in phrases and arguments in phrases (pattern 1), for arguments in phrases and arguments in clauses (pattern 2), for all three types of dependency relations (pattern 3) or for neither (pattern 4), but I do not expect languages to use a single coding strategy for dependency relations that are both semantically distinct (in terms of the modifier/argument distinction) and syntactically distinct (in terms of their phrasal vs. clausal nature), as in alignment patterns of type 5.

Before turning to the results of testing the hypothesis, I present the methodological background to the study in the next section.

⁸ Importantly, formal neutralizations need not affect each dependency relation from a given category; this is further discussed in Section 4.4.5.

4.4 Methodological background

4.4.1 The language sample

For this study, I composed a worldwide sample of 39 languages that is both genetically and geographically balanced. Maximal genetic independence is obtained by applying the Diversity Value technique proposed by Rijkhoff et al. (1993) and Rijkhoff & Bakker (1998) to Ruhlen's (1991) classification of the world's languages.⁹ This technique involves the computation of so-called 'Diversity Values' (henceforth DVs), which determine the number of languages to be selected from each family and subfamily on the basis of its internal complexity, given a particular desired sample size. The degree of internal complexity of a given (sub)family is based on the number of branches it contains (the width of the tree) and the time-frame over which the languages in these branches developed (the depth of the tree). Crucially, more internally complex (sub)families are assumed to show higher degrees of linguistic diversity than less internally complex (sub)families, thus yielding a higher DV. Language (sub)families with a higher DV are in turn represented by a higher number of languages in the sample than language (sub)families with a lower DV. Isolates are, by definition, part of any sample.

The genetically balanced sample is combined with a geographic stratification: within the restrictions of the DV technique, I selected solely languages spoken in non-contiguous areas. Finally, bibliographical restrictions apply: where the genetic and geographic stratifications allowed for a choice between languages from a given subgroup, selection was based on the quality of source materials and the availability of language informants.

⁹ Although Ruhlen's classification is largely outdated – given its postulation of some large, poorly supported language families like Amerindian (e.g. Kaufman 1990; Nichols 1990; Campbell & Poser 2008) – it is suitable for creating relatively small, but well-balanced, language samples, because in the DV technique a minimal sample consists of one language per first-order language family. Classifications such as Grimes (2000), used in *WALS*, and *Glottolog* (Hammarström et al. 2016) are more up-to-date, but also much subtler (*Glottolog*, for instance, distinguishes 433 first-order language groups), and are thereby more difficult to use in composing small, well-balanced samples with the DV technique. Also note that the DV technique requires internally complex (sub)families to be represented by a large number of languages, which partly compensates for relying on a rougher classification like Ruhlen's.

The 39-language sample is presented in Table 2 below. The first three columns present the number of languages selected from each family and subfamily, between brackets; language isolates are abbreviated as 'LI'. The fourth column presents the languages selected, and the final column lists the highest-order family of each sample language in the WALS (see footnote 9). Three languages (Etruscan, Meriotic and Nahali) selected by the sampling method were removed from the sample due to insufficient data.

Table 3: The 39-language sample

Language family (Ruhlen 1991)	Subfamilies	Language selected	Language family (WALS)	
Afro-Asiatic (2)	Chadic (1)	Hausa	Afro-Asiatic	
	Semitic (1)	Maltese	Afro-Asiatic	
Altaic (1)		Turkish	Altaic	
Amerind (5)	Northern (1)	Koasati	Muskogean	
	Andean (1)	Urarina	isolate	
	Eq.-Tucanoan (1)	Tariana	Arawakan	
	Ge-Pano-Carib (1)	Macushi	Cariban	
	Chib.-Paezan (1)	Sanuma	Yanomam	
Australian (2)	Pama-Nyungan (1)	Nyangumarda	Pama-Nyungan	
	Other (1)	Mangarayi	Mangarrayi-Maran	
Austriac (4)	Austro-Tai (2)	Daic (1)	Thai	Tai-Kadai
		Austronesian (1)	Drehu	Austronesian
	Austroasiatic (1)	Kharia	Austro-Asiatic	
	Miao-Yao (1)	Hmong Njua	Hmong-Mien	
Basque (LI)		Basque	isolate	
Burushaski (LI)		Burushaski	isolate	
Caucasian (1)		Abkhaz	Northwest Caucasian	
Chukchi-Kamchatkan (1)		Itelmen	Chukotko-Kamchatkan	
Elamo-Dravidian (1)		Tamil	Dravidian	
Eskimo-Aleut (1)		West Greenlandic	Eskimo-Aleut	

Table 3: The 39-language sample

Language family (Ruhlen 1991)	Subfamilies	Language selected	Language family (WALS)
Etruscan (LI)		–	–
Hurrian (LI)		Hurrian	–
Indo-Hittite (1)		English	Indo-European
Indo-Pacific (3)	Trans New Guinea (1)	Inanwatan	isolate
	Sko (1)	Skou	Sko
	West Papuan (1)	Tidore	West Papuan
Kartvelian (1)		Georgian	Kartvelian
Ket (LI)		Yeniseian	isolate
Khoisan (1)		Nama	Khoe-Kwadi
Korean-Japanese-Ainu (1)		Japanese	Japanese
Meroitic (LI)		–	–
Na-Dene (1)		Navajo	Na-Dene
Nahali (LI)		–	isolate
Niger-Kordofanian (3)	Niger-Congo (2)	N.-C. Proper (1)	Babungo
			Niger-Congo
		Mande (1)	Bambara
			Mande
		Kordofanian (1)	Krongo
			Kadu
			Lango
			Eastern Sudanic
Nivkh (LI)		Nivkh	isolate
Pidgins and Creoles (1)		Berbice Dutch Creole	–
Sino-Tibetan (1)		Mandarin Chinese	Sino-Tibetan
Sumerian (LI)		Sumerian	isolate
Uralic-Yukaghir (1)		Hungarian	Uralic

4.4.2 The dependency relations studied

In each sample language I investigated the coding of seven unique types of dependency relations, which belong to three types of constituents: possessive noun phrases, adpositional phrases and verbal main clauses containing one- and

two-place predicates. These constituent types were selected, because they are all in principle open to three types of coding strategies: flagging, indexing and the absence of both, i.e. zero-coding. As such, they are an interesting testing ground for studying the typological alignment of phrases and clauses.

Given my focus on phrasal and clausal dependency relations, I exclude dependent parts of compounds or incorporation constructions from the study. I also exclude dependents that take the form of (deranked) dependent clauses, such as relative clauses and complement clauses, as well as action nominalizations, participle and converb constructions. The latter typically show ‘mixed’ phrasal (i.e. nominal) and clausal (i.e. verbal) behavior, which makes it difficult to reliably classify them as either phrasal or clausal. Furthermore, I do not consider dependents that are internally complex, such as possessors in possessor chains. Finally, three-participant predicates are left out of present consideration.

Table 4 below presents the seven dependency relations studied, in terms of their combined semantic and syntactic type, following the five-way typology given in Section 4.3 (in the first column), their constituent type (in the second column), the type of dependent (the third column) and the type of head (the fourth column), followed by some illustrative examples (in the final column).

Table 4: Inventory of the dependency relations studied

Semantico-syntactic category	Constituent type	Dependent	Head	Example
Modifiers in phrases	Possessive NP	Possessor	Concrete, inanimate noun	<i>John’s book</i>
Arguments in phrases	Possessive NP	Possessor	Kinship noun or body part noun	<i>Mary’s hair</i>
	Adpositional phrase	Adpositional argument	Adposition	<i>Near the girl</i> <i>Inside (of) the house</i>
Arguments in clauses	Main clause	S _A	Verb	<i>The dog runs.</i> <i>John swims.</i>
		S _P		<i>The grandfather dies.</i> <i>The cat falls.</i>
		A		<i>Mary hits John.</i>
		P		<i>The cat kills the mice.</i>

The overview of dependency relations in Table 4 requires some explanation. First, let us consider the way in which modifiers and arguments are distinguished in this paper. As already noted, this distinction is a notoriously difficult one to draw. On the one hand, the valency of individual lexical items, be they nouns or verbs, displays significant cross-linguistic variation. With respect to nouns, it is often observed that languages with a formal distinction between alienable and inalienable possession vary in the types of nouns participating in each class: while in some languages only a set of kinship terms or body part terms is inalienable, in other languages this class additionally includes spatial nouns, (other) parts of wholes, property nouns and a set of culturally-specific objects (such as ‘net’ or ‘hammock’) (Nichols 1988: 572; Nichols & Bickel 2013b; Chappell & McGregor 1996: 8–9).

The same type of variation is attested with verbs: while the verb *like* in English takes a subject and a direct object, as shown in (10), the German equivalent *gefallen* takes an indirect object and a subject respectively, as shown in (11):

(10) *I like him.*

(11) *Er gefällt mir.*
 He.NOM like.3SG.PRS I.DAT
 ‘I like him.’

Such cross-linguistic variation in ‘predicate classes’ (also called ‘verb classes’ or ‘valency classes’) has been widely studied, a recent example being the Leipzig Valency Patterns project (Hartmann et al. 2013).

On the other hand, most tests used to distinguish modifiers and arguments are language-specific and are therefore not applicable cross-linguistically (cf. Haspelmath 2014; Forker 2014). For instance, arguments are often defined as being obligatorily expressed, but in many languages all sorts of dependents may be freely dropped – a phenomenon recently studied under the name of *referential density* (Bickel 2003). Notably, this problem is not only known from research into clausal alignment and verbal valency (cf. Witzlack-Makarevich 2011: 99–137 and Haspelmath & Hartmann 2015) but from studies on the valency of nouns as well (cf. Partee & Borschev 2003). In order to reliably distinguish modifiers and arguments across the languages of the sample, I therefore focus on prototypical instances of such dependents only. Prototypical relational (argument-taking) possessums are kinship terms and body part terms, while prototypical non-relational (modifier-taking) possessums are concrete, inanimate objects, such as ‘pot’ or ‘pen’. Across individual languages,

inalienable coding is minimally used for (a subset of) kinship terms and/or body part terms, while concrete object nouns are typically found in the alienable class (cf. Nichols 1988: 572; Nichols 1992: 120; Haiman 1985: 136). Moreover, as shown in Section 4.2, such nouns clearly qualify as, respectively, relational and non-relational on the basis of their inherent semantic properties.¹⁰

The same approach is adopted for the delineation of verbal arguments. With respect to two-participant verbs, I follow work on main clause alignment (such as Comrie 1989: 111, 2013 and Lazard 2002) in focusing solely on prototypical action verbs, i.e. physical effect verbs like ‘hit’, ‘break’ and ‘kill’.¹¹ Such verbs are semantically prototypical two-participant predicates, and they are at the core of transitive verb classes in individual languages (cf. Haspelmath 2014, and references therein). With respect to one-participant verbs, I focus on two generalized predicate classes: (i) basic motion and activity verbs such as ‘run’, ‘jump’ and ‘play’, representative of one-participant active verbs, and (ii) uncontrolled change of state verbs like ‘die’, ‘fall’ and ‘burn’, representative of one-participant stative verbs.¹² In this way, I accommodate languages with so-called ‘split-S’ or ‘stative/active’ systems of clausal alignment, in which Ss are treated either like the A or like the P, depending on semantic factors, such as affectedness and control (cf. Mithun 1991b; Donohue & Wichmann 2008).

Note that the semantic classes of nouns and verbs selected for this paper need not receive internally consistent coding. In Koasati (Muskogean, United

¹⁰ Possessors of kinship terms and body part terms are by definition human or animate; inanimate possessors are found only with spatial nouns (*the inside of the house*) and other part-whole relations (*the wheels of the bicycle*). As a result, animacy-based splits in possessive coding, and in alignment in general, are not included in this study. This is unproblematic as there is only a single sample language, Burushaski, in which animacy conditions alignment (specifically, the coding of Ss and Ps). Importantly, none of the coding options for possessed nouns other than those currently investigated contradict the main findings of the study.

¹¹ An alternative approach is to focus on the major two-participant predicate class (in the sense of lexical types) in any given language (cf. Bickel et al. 2015; Witzlack-Makarevich 2011: 111–114). I did not adopt this approach because counts are rarely reported in the relevant sources, or have not been made at all. Moreover, as pointed out by Haspelmath (2015), both approaches probably yield the same results, as prototypical action verbs are typically part of the largest class of two-participant predicates across individual languages. My data support this observation.

¹² I did not consider other prototypical stative verbs, i.e. properties such as ‘hungry’, ‘black’ and ‘little’, since these concepts are often treated as adjectives rather than as verbs in individual languages (as in English, for instance). In some sample languages, only such properties are stative verbs, which therefore do not surface in the data.

States), for instance, inalienable prefixes are used for only a minority of kinship and body part terms; the majority of such terms receive alienable prefixes (Kimball 1991: 433). In the case of such lexical splits in possessive coding, I take both coding strategies into account, i.e. argument-taking nouns are treated as allowing two types of possessive coding.¹³

The same kind of lexical variation is sometimes attested within the three classes of verbal predicates considered: in certain varieties of the isolate Basque (France, Spain), for instance, arguments of *korritu* ‘run’ receive ergative coding, while absolutive coding is more regularly used for arguments of *dantzatu* ‘dance’, *saltatu* ‘jump’ and *hegatu* ‘fly’ (Aldai 2008: 203). Similarly, in Koasati, most two-participant action verbs, including *batáplin/bóklin* ‘hit’ and *wakóʔin/wakó:lin* ‘break’ are part of the same predicate class, while *latáʔkan/látkan* ‘kick’ and *ciʔipkan/ciʔkan* ‘spear’ are part of another (Kimball 1991: 56–89).¹⁴ In such cases, I adopt a lexical-typological approach (cf. Nichols et al. 2004; Nichols 2008; Haspelmath 2013c, 2015), i.e. I focus on the coding of the translational equivalents of individual lexical items, more specifically ‘run’ (for one-participant active verbs), ‘die’ (for one-participant stative verbs) and ‘hit’ (for two-participant verbs).¹⁵ The choice for these particular lexemes is mainly a practical one: information on their coding is commonly given in language descriptions and is easily checked with language informants. Also, they can take arguments with different referential properties, e.g. human, animate, definite, indefinite, which is important given the inclusion in this study of such properties as conditions on alignment. In one case, namely Skou (Skou, Indonesia; Donohue 2004: 214–215), the indexing of ‘hit’ turned out to be highly irregular, both in terms of form (stem suppletion and special initial consonant alternations) and frequency (few other verbs show this pattern) and is therefore not representative of its class; I therefore adopted the regular, more frequent pattern of indexing for As in this language.

¹³ In many languages, inalienable nouns may take alienable coding but only when possessed in an alienable (i.e. non-relational) manner; since such differential coding goes hand in hand with a semantic shift, it is left out of present consideration.

¹⁴ As in other Muskogean languages, pairs of Koasati verb forms mark the singularity vs. plurality of the subject or object, or the number of times the action is performed.

¹⁵ This approach could not be straightforwardly applied to (modifier-taking and argument-taking) possessum nouns and adpositions, for two reasons. First, there was no lexical item for which the selected coding strategy was available in all 39 sample languages. Second, I had no strong intuitions, nor found any suggestions in the typological literature, on the most prototypical instances of such items.

Additionally, differential coding in clauses may be conditioned by a range of other factors, such as clause type (main vs. dependent clauses), polarity (positive vs. negative clauses), tense/aspect/mood and diathesis. Following much other typological work (e.g. Siewierska 2013b; Comrie 2013) I focus solely on main, positive, declarative clauses in the indicative mood, thereby excluding most of those factors from the study. Only tense/aspect-based splits in alignment have been taken into account, in the same way as the lexical splits in possessive coding discussed above (and referential splits, discussed in the next subsection). Overtly derived diathesis alternations, such as passive or antipassive constructions, have not been taken into account; one sample language, Hurrian (Uralic, Turkey), has a formally underived antipassive construction, yielding three additional alignment patterns (one for indexing and two for flagging). Finally, clausal alignment may depend on so-called ‘co-argument’ or ‘scenario’ conditions (cf. Witzlack-Makarevich 2011: 94; Witzlack-Makarevich et al. 2016), i.e. the coding of A and P depends on the referential properties of either or both arguments. This is the case in Mangarayi (Mangarrayi-Maran, Australia), where identity of the A prefixes with the S prefixes depends on the person/number/gender properties of both A and P (Merlan 1982: 160). Each combination of A and P properties yielding an alignment split has been taken into account, in the same way as referential conditions, described in the next subsection.

Focusing in turn on adpositional phrases, I follow Bakker (2013) in defining adpositions as formally independent elements that display morphosyntactic behavior distinct from regular verbal, nominal and adverbial elements in any given language. The first part of the definition excludes case affixes and adpositional clitics that attach to NPs rather than to nouns; although in many languages such markers historically derive from freestanding adpositions, they now attach to the argument, rather than functioning as heads of phrases themselves. As a consequence, they cannot take indexes or govern flagging of the argument, like freestanding adpositions, and are therefore irrelevant to the present paper. The second part of the definition excludes adpositional meanings expressed by nouns, verbs or adverbs. Typical ‘nominal’ behavior includes the ability to take (the full range of) case markers or other means of flagging available in the language, and the possibility of being pluralized, lexically modified or marked as (in)definite. Typical ‘verbal’ behavior includes the use of tense/aspect/mood markers and the ability to be modified by adverbs. Typical ‘adverbial’ behavior includes an item’s ability to be separated from or used without an argument, and the absence of case assignment to the argument. I define adpositions as items that differ from nouns, verbs or adverbs

in at least one of these properties, or that can no longer function as these parts of speech at all.¹⁶ Seven sample languages (Inanwatan, Itelmen, Ket, Mangarayi, Nyangumarda, Thai, West Greenlandic) do not have any adpositions according to this definition, or the respective language sources lack sufficient information to permit determination of the presence of such a class.¹⁷ Finally, adpositions commonly show lexically-conditioned splits in coding: in some languages, only a subset of adpositions can be indexed or take a particular case marker. As with possessum nouns, each coding option is taken into account in individual languages.

In addition to the factors discussed in this subsection, another major source of coding splits is variation in the referential properties of modifiers and arguments. In the next subsection I give examples of cross-linguistically attested splits, and I explain how they are treated in the present study.

4.4.3 Referential conditions on alignment and their methodological treatment

With respect to clausal alignment, typologists have recognized that “a characterization of whole languages as representing a particular alignment pattern in case marking, agreement, etc. is known to provide a simplified picture as languages can show variations in their alignment patterns often referred to as alignment splits” (Witzlack-Makarevich 2011: 65). Alignment splits are often conditioned by referential properties of dependents, such as person, animacy, humanness, definiteness, specificity and anaphoricity, i.e. the (pro)nominal nature of the dependent. This phenomenon has been widely studied in the clausal domain under a variety of labels, including ‘differential subject/object marking’ (Bossong 1985, 1998; Comrie 1989), ‘differential argument marking’ (Baerman 2008) and ‘split ergativity’ (Silverstein 1976). However, referentially-

¹⁶ The motivation for this relatively ‘tolerant’ definition of adpositions is mainly practical: most sources provide insufficient information to allow determination of the full extent of the deranked behavior of such items in individual languages.

¹⁷ A distinction is sometimes made in the literature between lexical and grammatical adpositions: the former are instantiated by lexical items, and are thus able to head an adpositional phrase, while the latter solely express semantic functions (cf. Hengeveld & Mackenzie 2008: 251, 401). However, given a lack of consensus on the defining properties of the two types of adpositions (cf. Keizer 2008; Mackenzie 2013), such a distinction is not made here.

conditioned splits in coding are also found in the phrasal domain, where they remain understudied. A number of examples are provided below.

Starting with flagging, it may be observed that pronominal possessors are often expressed differently from nominal possessors; most Indo-European languages, such as English, have unique sets of possessive pronouns. Anaphoricity is also relevant for possessive flagging in Krongo (Kadu, Sudan): while pronominal possessors of kinship terms are marked with the possessive prefix *kà-* (12a), nominal possessors of kinship terms may be marked with the genitive prefix *má-* (12b):

- (12) a. *ní-imò kà-tí*
 mother POSS-1SG
 'my mother'
- b. *ní-imò má-Kàkká*
 mother GEN-Kaka
 'Kaka's mother' (Reh 1985: 318, 316)

Another type of referential split in possessive flagging is found in Drehu (Austronesian, New Caledonia), where the marker *n(e)* is used for common nouns (13a), while the marker *i* is used for proper nouns (13b):

- (13) a. *la uma ne la qatreföe*
 ART house ASS ART old.woman
 'the house of the old woman'
- b. *la uma i Hagee*
 ART house ASS Hagee
 'the house of Hagee' (Moyse-Faurie 1983: 59)

Referential factors may also condition the presence vs. absence of means of flagging: cross-linguistically, highly referential nouns are often overtly marked, while low referential nouns are not (cf. Chappell & McGregor 1989; Aikhenvald & Dixon 2013: 44). A well-studied case in point is Turkish (Altaic, Turkey), where the genitive case suffix is used on specific nouns (14a), but not on non-specific nouns (14b):

- (14) a. *bir balığ-m kuyruğ-u*
 a fish-GEN tail-3SG.POSS
 'the tail of a fish'

- b. *bir balık kuyruğ-u*
 a fish tail-3SG.POSS
 ‘a fish tail’ (Nilsson 1985: 69)

Differential flagging in adpositional phrases is also attested. Consider, again, Turkish, where a subset of postpositions assigns the genitive case to pronouns (except for 3rd person plural pronouns), as shown in (15a), but not to nouns (15b):

- (15) a. *on-un için*
 he-GEN for
 ‘for him’
 b. *Ahmet için*
 Ahmet for
 ‘for Ahmet’ (Kornfilt 1997: 301)

In Kharia (Austro-Asiatic, India), postpositional arguments prefer the genitive case marker =*yaʔ* when definite (16a), but the absence of case-marking when indefinite (16b):

- (16) a. *aba=ɖom=aʔ lekhe=ga*
 father=3POSS=GEN lekhe=FOC
 ‘like his father (he spread out the net)’
 b. *samudar lekhe*
 ocean like
 ‘like an ocean’ (Peterson 2011: 204, 205)

Referential properties also commonly condition indexing: specifically, Siewierska (2004: 148–162) found that highly referential dependents are more likely to be indexed than low-referential dependents, in clauses but also in phrases. Anaphoricity is a common referential condition on indexing: languages often use indexes only in the absence of a nominal dependent in the phrase.¹⁸ A

¹⁸ In some languages, the index may nevertheless co-occur with a free pronoun, while in others this is impossible as well. The inverse pattern, i.e. the presence of indexes with nouns but the absence of indexes with free pronouns, is cross-linguistically very rare, and is not attested in my sample (cf. Siewierska 2004: 154 for an example). For this reason, in this study I only consider the co-occurrence of indexes with nominal dependents, not with pronominal ones.

case in point is Lango (Eastern Sudanic, Uganda), where this is found in possessive NPs, as shown in (17a) and (17b), and in adpositional phrases, as shown in (17c) and (17d):

- (17) a. *wì rwòt*
 head king
 'the king's head'
- b. *wì-è*
 head-3SG
 'his/her head'
- c. *bòt dákô*
 to woman
 'to the woman'
- d. *bòt-é*
 to-3SG
 'to him/her' (Noonan 1992: 157, 82, 42, 107)

Another common factor governing the use of indexes is person. In Drehu, as shown in (18a), 1st person possessors are indexed with a suffix, while, as shown in (18b), 2nd and 3rd person possessors take the form of free pronouns plus the marker *i*, also used for proper nouns (cf. (13b)):

- (18) a. *uma-ng*
 house-1SG
 'my house'
- b. *la uma i angeic*
 ART house ASS he
 'his house' (Moyse-Faurie 1983: 61, 59)

As with flagging, referential factors may condition not only the presence vs. absence of indexes, but also the type of index-set selected. In Lango, for instance, singular (pronominal) possessive modifiers take a unique possessive index-set, as shown in (19a), while plural (pronominal) possessive modifiers take the index-set additionally used for prepositional arguments and P arguments, as demonstrated in (19b), (19c) and (19d) respectively:

- (19) a. *gúlú-mérê*
 pot-3SG.POSS
 'his pot'

- b. *pàlà-gí*
knife-3PL.ACC
'their knife'
- c. *tê-gí*
under-3PL.ACC
'under them'
- d. *ò-nènò-gí*
3SG.NOM-see.PRF-3PL.ACC
'He saw them.' (Noonan 1992: 79, 79, 107, 37)

In some languages, the same referential factor(s) are relevant in phrases and in clauses. In Kharia, for instance, definiteness not only conditions the use of the genitive case on postpositional arguments, as shown in (16), but also the use of the oblique marker =*te* on Ps in main clauses, as shown in (20a) and (20b):

- (20) a. *dáʔ=te* *yo=yoʔ*
water=OBL see=ACT.PST
'(He) saw the water.'
- b. *ɪn* *dáʔ* *biʔh=oʔ*
1SG water pour.out=ACT.PST.1SG
'I poured water out.' (Peterson 2011: 144, 265, p.c.)

Similarly, indexing in Lango is sensitive to anaphoricity not only in possessive NPs and prepositional phrases, as shown in (17), but also for Ps in main clauses, shown in (21a) and (21b):

- (21) a. *dákó* *ò-jwátò* *lócà*
woman 3SG-hit.PRF man
'The woman hit the man.'
- b. *dákó* *ò-jwát-é*
woman 3SG-hit.PRF-3SG
'The woman hit him.' (Noonan 1992: 128, 239)

Crucially, referential splits in the coding of dependency relations, be they phrasal or clausal, may yield splits in the overall alignment of a given language. In this study, I take into account, per sample language, the set of dependency relations covered by the same coding strategy and the type of (referential and/or lexical) condition under which this occurs, following practice in recent work on clausal alignment typology (Witzlack-Makarevich 2011; Bickel et al. 2013; Bickel

et al. 2015). Each alignment pattern is then classified in terms of the five-way typology presented in Section 4.3. A number of concrete examples of how this works are provided below.

Mapping the alignment patterns in a given language is relatively straightforward when only a single referential condition splits coding in one (or more) dependency relations. In Macushi (Cariban; Brazil, Guyana, Venezuela), for instance, indexing is conditioned by anaphoricity in all seven dependency relations studied: as in Lango, indexes are used in the absence of a nominal dependent ([-N]), but not in the presence of a nominal dependent ([+N]). This split yields two unique alignment patterns, presented in Table 5. The table presents the coding of individual dependency relations (in the columns) under each referential split (in the final column). Formally identical index-sets are given the same abbreviation.

Table 5: Indexing alignment in Macushi

Possessive modifier	Possessive argument	Adpositional argument	S _A	S _P	A	P	Referential conditions
suffix	suffix	∅	∅	∅	∅	∅	[+N]
ACC prefix	ACC prefix	ACC prefix	NOM prefix	NOM prefix	ERG suffix	ACC prefix	[-N]

A split into two alignment patterns is impossible when multiple dependency relations display differential coding, under varying referential conditions. In Lango, exemplified in (17), (19) and (21), indexing is conditioned by anaphoricity (for possessors, adpositional arguments and Ps), number (for possessors) and humanness (for Ps). In order to map the alignment variation in the language, these referential properties have to be cross-classified, which yields four unique alignment patterns, given in Table 6. Importantly, splits in coding may but need not trigger splits in alignment. As shown in the second row of the table, possessors, adpositional arguments and Ps alternate between zero-indexing (for nouns) and accusative indexing (for human, plural pronouns). This is a form of differential coding, but not of differential alignment, since, under both conditions, the relevant dependency relations are covered by a single (differential) coding strategy. Both conditions therefore take up only a single row in the table.¹⁹

¹⁹ Note additionally that splits in alignment, as studied in this paper, do not subsume splits in clausal alignment. The differential indexing of Ps in Lango governs an overall alignment split, but clausal alignment unequivocally follows an accusative pattern.

Table 6: Indexing alignment in Lango

Possessive modifier	Possessive argument	Adpositional argument	S _A	S _P	A	P	Referential conditions
∅	∅	∅	NOM	NOM	NOM	∅	[+N]
ACC suffix	ACC suffix	ACC suffix	prefix	prefix	prefix	ACC suffix	[-N],[+hum],[-sg]
ACC suffix	ACC suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	∅	[-N],[-hum],[-sg]
AL suffix	INAL suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	ACC suffix	[-N],[+hum],[+sg]
AL suffix	INAL suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	∅	[-N],[-hum],[+sg]

Alignment splits may also be conditioned by lexical factors; as discussed in the previous subsection, such factors are taken into account when relevant in (possessive and adpositional) phrases. Lexical alignment splits are treated in the same way as referential splits: as separate subsystems in individual languages. For instance, in the isolate Burushaski (Pakistan), different postpositions trigger four different case suffixes on the argument: absolutive, relative, dative and ablative. This yields a three-way split in alignment, as shown in Table 7 below. The alternation between the dative and ablative case does not yield an alignment split, as these case markers are not used in any of the other dependency relations studied.

Table 7: Flagging alignment in Burushaski (for inflection classes 1, 3 and 4)

Possessive modifier	Possessive argument	Adpositional argument	S _A	S _P	A	P	Lexical conditions
REL <i>-e</i>	REL <i>-e</i>	ABS ∅	ABS ∅	ABS ∅	REL <i>-e</i>	ABS ∅	Postposition class 1
REL <i>-e</i>	REL <i>-e</i>	REL <i>-e</i>	ABS ∅	ABS ∅	REL <i>-e</i>	ABS ∅	Postposition class 2
REL <i>-e</i>	REL <i>-e</i>	DAT <i>-(a)r</i> ABL <i>-cum</i>	ABS ∅	ABS ∅	REL <i>-e</i>	ABS ∅	Postposition class 3

It is quite common for alignment to be governed by lexical *and* referential factors. In Burushaski, a single relative case is used on possessors and As of all

inflection classes except one: human females.²⁰ This inflection class formally distinguishes a genitive case (on possessors) and an ergative case (on As). As a result, three additional alignment patterns have to be taken into account for this inflection class, given in Table 8. As shown in Tables 7 and 8, each combination of the lexical and referential conditions relevant in Burushaski yields a unique alignment pattern.²¹

Table 8: Flagging alignment in Burushaski (for inflection class 2)

Possessive modifier	Possessive argument	Adpositional argument	S _A	S _P	A	P	Lexical conditions
GEN <i>-mu</i>	GEN <i>-mu</i>	ABS \emptyset	ABS \emptyset	ABS \emptyset	ERG <i>-e</i>	ABS \emptyset	Postposition class 1
GEN <i>-mu</i>	GEN <i>-mu</i>	GEN <i>-mu</i>	ABS \emptyset	ABS \emptyset	ERG <i>-e</i>	ABS \emptyset	Postposition class 2
GEN <i>-mu</i>	GEN <i>-mu</i>	DAT <i>-(a)r</i> ABL <i>-cum</i>	ABS \emptyset	ABS \emptyset	ERG <i>-e</i>	ABS \emptyset	Postposition class 3

Finally, lexical and referential factors may interact, such that lexical factors give rise to additional alignment patterns for a subset of referential dependents only. This is the case in Lango, where a small number of kinship terms take the alienable index-set, rather than the inalienable index-set (Noonan 1992: 82). Since both sets are formally distinct only for singular pronominal possessors (cf. Table 6) this lexical condition yields only two additional alignment patterns, presented in Table 9.

²⁰ In Burushaski, class assignment is semantic and can therefore be described in terms of referential properties (human male vs. human female vs. animate vs. inanimate); however, this is not the case in all sample languages (and even in Burushaski there are some lexical exceptions: some inanimates form part of the ‘animate’ inflection class). Therefore, alignment splits based on inflectional classes are analyzed in the same way as referentially-conditioned and lexically-conditioned splits.

²¹ Flagging alignment in Burushaski additionally exhibits a tense-based split – A arguments receive the absolutive case in the future – which, for convenience sake, has not been illustrated here.

Table 9: Indexing alignment in Lango for a small subset of kinship terms

Possessive modifier	Possessive argument	Adpositional argument	S _A	S _P	A	P	Referential conditions
AL suffix	AL suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	ACC suffix	[-N],[+hum],[+sg]
AL suffix	AL suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	∅	[-N],[-hum],[+sg]

In sum, this subsection has shown how referential factors may impact the coding of possessive NPs and adpositional phrases in individual languages. More importantly, I have shown how referential and lexical factors can be taken into account in comparing languages in terms of the alignment of phrases and clauses. In the next subsection, I discuss further issues in data analysis and coding. The classification of alignment patterns following the five-way typology in Section 4.3 is discussed in Section 4.4.5.

4.4.4 Further methodological preliminaries

The alignment of flagging and the alignment of indexing are considered separately in this study. Free and clitic person forms are considered to be pronouns – and thus treated under the alignment of flagging – while affixal person forms are considered to be indexes – and thus treated under the alignment of indexing.²² As is well known, pronouns need not be obligatorily expressed. Following Comrie (2013), I take the alignment of pronouns into account irrespective of conditions on their use; only when they cannot be used in dependent positions at all, or no information on their use is available, are they left out of present consideration.

Within the alignment of flagging, a further distinction is made between flagging of the dependent and flagging of the head. This is done to accommodate languages in which the dependent and the head are both marked, but with distinct means of flagging. In West-Greenlandic (Eskimo-Aleut, Greenland; Fortescue 1984: 172, 216), for instance, possessive modifiers and possessive arguments both receive a genitive case marker – and are thus considered under the alignment of dependents – but only possessive modifiers are additionally flagged on the head with the suffix *-uti* – which are thus

²² In most sample languages, clitics and affixes could be reliably distinguished; in only a few cases did a lack of information force me to follow the orthography of the source consulted.

considered under the alignment of heads. Similarly, the alignment of free pronouns is considered separately from the alignment of cliticized pronouns on the head. In a few cases, formal identities between coding strategies do not show up in the data, because they differ in locus (Nichols 1986, 1992). For instance, in Skou (Donohue 2004: 190–191), a single set of pronominal enclitics is used on possessums – head-marking, thus considered under the alignment of heads – and on A arguments – dependent-marking, thus considered under the alignment of dependents. A dependent or head may also be marked twice, i.e. by a combination of prefixes and suffixes. This applies to case-marking in Mangarayi (Merlan 1982: 56) and the indexing of Ss in Itelmen (Bobaljik & Wurmbbrand 2002). In such cases, I treat the combination of markers as a single coding strategy. Finally, indexes that differ in their position with respect to the head, i.e. prefixes vs. suffixes, are considered to be separate indexing strategies, irrespective of identities in form.

The lack of indexes is interpreted as the absence of indexing rather than as indexing realized by zero. The same applies to flagging: case marking with zero exponence (typically nominative and absolutive) is analyzed on a par with the absence of other means of flagging (such as differential genitive or accusative case marking). This is motivated by the present approach to alignment patterns as transcending the domain of the clause, under which a zero nominative or absolutive case form serves the same function as, for instance, a zero genitive case form: in both cases, a given dependency relation receives no overt form of coding.

Finally, allomorphic sets of indexes or case forms are considered to be formally identical when conditioned by the same factor, typically the same morpho-phonological rule. Sometimes, conditions are not logically applicable across dependency relations, e.g. when the choice of an indexical allomorph depends on tense/aspect/mood with verbs, but on lexical noun class with possessum nouns. Since it is impossible to empirically determine the formal identity of the allomorphs in such cases, I take the unitary approach of considering both sets to be identical.

4.4.5 Typological classification of alignment patterns

Each alignment pattern attested in the sample is assigned to one of the five types in the typology presented in Section 4.3, repeated for convenience in Table 10:

Table 10: Typology of alignment types

	1	2	3	4	5
Modifier in phrase	A	A	A	A	A
Argument in phrase		B		B	B
Argument in clause	B			C	A

The classification of language-specific alignment patterns following this typology is exemplified for the inventory of indexing alignment patterns in Lango, re-presented in Table 11:

Table 11: Indexing alignment in Lango

Modifier in phrase	Argument in phrase		Argument in clause				Referential conditions	Type
Possessive modifier	Possessive argument	Adpositional argument	S _A	S _P	A	P		
∅	∅	∅	NOM	NOM	NOM	∅	[+N]	3
ACC suffix	ACC suffix	ACC suffix	prefix	prefix	prefix	ACC suffix	[-N],[+hum], [-sg]	
ACC suffix	ACC suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	∅	[-N],[-hum], [-sg]	1
AL suffix	INAL suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	ACC suffix	[-N],[+hum], [+sg]	2
AL suffix	INAL suffix	ACC suffix	NOM prefix	NOM prefix	NOM prefix	∅	[-N],[-hum], [+sg]	4

The first alignment pattern in the table is a type 3 pattern, because dependency relations from each of the three categories are formally neutralized: possessive modifiers (modifiers in phrases), possessive and adpositional arguments (arguments in phrases) and Ps (arguments in clauses) may each be expressed by means of the accusative suffix or remain unindexed. The second alignment pattern in the table is a type 1 pattern, since a single formal treatment is given to phrases (possessive modifiers, possessive arguments and adpositional arguments, expressed by the accusative suffix), while another is given to clauses (Ss and As, expressed by the nominative prefix, and Ps, which remain unindexed). The third pattern belongs to type 2, since possessive modifiers (modifiers in phrases, expressed by the alienable suffix) are given a separate treatment from phrasal and clausal arguments (adpositional phrases and Ps, expressed by the accusative suffix). The final pattern belongs to type 4, as it

shows the complete absence of cross-categorical neutralizations – although neutralizations within each category are, of course, possible, such as that of S_A , S_P and A , expressed by the nominative prefix.

Importantly, the neutralizations defining a pattern's type need not affect each dependency relation from a given category: in Lango, S_{AS} , S_{PS} and A_S do not participate in the type 3 pattern (full neutralization), while in the type 2 pattern (modifiers vs. arguments) only adpositional phrases and P_S participate. Also note that assignment of a pattern to one of the five types is always based on the most extensive type of cross-categorical neutralization, i.e. the coding strategy that covers the most categories. The first alignment pattern in Table 11, for instance, is classified as a type 3 pattern, because the zero and accusative indexes cover dependents from all three categories: phrase modifiers, phrase arguments, and clause arguments. This applies irrespective of additional formal neutralizations, e.g. if not only S_S and A_S , but also adpositional arguments are expressed with a nominative prefix.

Where both alignment patterns cover an equal number of dependency relations, the criterion of maximum cross-categorical neutralization cannot enforce a decision between one and the other alignment type. This applies to patterns of types 1 and 2, and those of types 2 and 5. Burushaski provides an example of the former, as shown in Table 12.

Table 12: Flagging alignment in Burushaski (for inflection class 2 and a subset of postpositions)

Possessive modifier	Possessive argument	Adpositional argument	S_A	S_P	A	P	Lexical conditions
GEN <i>-mu</i>	GEN <i>-mu</i>	ABS \emptyset	ABS \emptyset	ABS \emptyset	ERG <i>-e</i>	ABS \emptyset	Postposition class 1

On the one hand, this pattern shows the formal neutralization of phrases (possessive modifiers and possessive arguments, expressed by a single genitive case marker) which is characteristic of type 1 patterns, while on the other hand it shows the formal neutralization of arguments (adpositional and verbal arguments, expressed by the absence of flagging) which is characteristic of type 2 patterns. Since both neutralizations cover dependents of two categories, such patterns are considered a separate type, unambiguously named 'type 1&2' in the remainder of the paper. This brings the number of possible alignment types to six. Alignment patterns showing neutralizations of type 2 and the unpredicted type 5 are unequivocally classified as type 5 patterns, in order to maximize the *a*

priori likelihood of finding a counterexample in the data.²³

The seven dependency relations investigated in this study can be formally neutralized in 877 distinct ways in any given language. In other words, there are 877 unique alignment patterns logically available in each sample language.²⁴ The distribution of these 877 unique patterns over the six alignment types in the typology is presented in Table 13 below. As shown in the table, type 4 patterns have the lowest number of unique instantiations (30 patterns), which is a logical consequence of the fact that cross-categorical neutralizations (absent from patterns of this type) are given preference in the assignment of patterns to types. The distribution also reflects the fact that the seven dependency relations studied do not divide equally over the three categories, i.e. I study one phrasal modifier, two phrasal arguments and four clausal arguments. For instance, in a type 3 pattern, each of the two phrasal arguments and four clausal arguments can in principle participate (note that phrasal modifiers are always part of such a pattern), while in a Type 1 pattern only the single phrasal modifier and the two phrasal arguments can participate. The 877 unique alignment patterns are presented in Appendix III.

Table 13: Distribution of unique, logically possible alignment patterns over alignment types

Alignment type	Frequency of unique, logically possible alignment patterns
1	45
1 & 2	74
2	173
3	235
4	30
5	320
Total	877

The distribution of the six alignment patterns in the data is discussed in the next section, where the results of the study are presented. I will first focus on

²³ Note that the alternative approach – the introduction of a separate alignment type ‘2&5’ – does not affect the results of this study, since, as shown in the next section, type 5 neutralizations are cross-linguistically very rare.

²⁴ The number of different ways in which a set of variables can be partitioned is expressed by so-called ‘Bell numbers’; for 7 variables, this number is 877 (Guichard forthcoming: 21–27).

the alignment of flagging (Section 4.5.1) and then on the alignment of indexing (Section 4.5.2). An overview of the alignment patterns in each sample language is provided in Appendix IV.

4.5 Results

4.5.1 Alignment of flagging

Table 14 below presents the distribution of the six alignment types for flagging in the sample; it gives the alignment type (in the first column), the number of observed alignment patterns of each type (in the second column), the number of unique, logically possible patterns of each type (in the third column, repeating Table 13), the number of patterns to be expected in the data given the actual number of observations (in the final column), and the factor difference between the observed and expected frequencies of each type (in the final column). Note that the table counts alignment patterns (N=175), not sample languages (N=39), since any given language may exhibit multiple alignment patterns. The distribution of alignment types over the sample languages is presented in a separate table, given later in this section.

Table 14: The observed and expected distribution of alignment of flagging

Type	Observed frequency	Frequency of unique, logically possible alignment patterns	Expected frequency	Factor difference observed/expected frequencies
1	56	45	8.98	6.24
1 & 2	35	74	14.77	2.37
2	27	173	34.52	0.78
3	54	235	46.89	1.15
4	2	30	5.99	0.33
5	1	320	63.85	0.02
Total	175	877	175	-

As shown in the second column of this table, alignment patterns of type 5, characterized by the formal identity of (possessive) modifiers in phrases and (one or more) arguments in clauses to the exclusion of arguments in phrases, are cross-linguistically very rare: although this type has the highest number of unique, logically possible patterns (320 out of 877, i.e. 36%), it has only a single

attestation in the data. Type 4 patterns, characterized by the full formal differentiation of the three types of dependency relations, are also uncommon, but this is unsurprising given that only 3% (30 of 877) of all logically possible alignment patterns are of this type. The four other alignment types are attested much more frequently. These data confirm the main hypothesis for flagging: if the same flagging strategy is used for modifiers in phrases as for arguments in clauses, arguments in phrases receive this flagging strategy as well. Alignment patterns in which a single means of flagging is used for modifiers in phrases and for arguments in clauses, but not for arguments in phrases, are rare. This finding is captured by the scale in (9), repeated in (22):

(22) modifier in phrase – argument in phrase – argument in (main) clause

Flagging strategies, i.e. overt means of flagging or their absence, strongly tend to cover continuous segments of this scale. Languages may formally neutralize the modifier/argument distinction (as in type 1 neutralizations), the phrase/clause distinction (as in type 2 neutralizations), both (as in type 3 neutralizations) or neither (as in type 4 neutralizations), but the formal neutralization of dependency relations that are both semantically and syntactically distinct is strongly dispreferred (type 5 neutralizations). This finding demonstrates the typological relevance of the opposition between modifiers and arguments, and that between phrases and clauses in the formal expression of dependency relations through flagging.

Statistical support for this finding is provided by a chi-square goodness-of-fit test, which tests whether the distribution of alignment types observed in the data (in the second column) differs significantly from the distribution of alignment types to be expected in the data (in the fourth column) on the basis of the number of unique alignment patterns belonging to each type (in the third column). Since the distribution of unique alignment patterns over the different types is inherently skewed, it possibly interferes with the observed distribution of alignment types in the data; the chi-square test determines the likelihood of such interference. The expected distribution of alignment types is computed by multiplying the probability of each type with the total number of alignment patterns attested (N=175); for instance, 45 of the 877 unique patterns are of type 1, which, given the total sample size of 175 patterns, yields an expected number of 9 patterns of this type. The test demonstrates a statistically significant difference between the two distributions ($\chi^2 = 341.19$, $p < 0.0001$), showing that it is highly unlikely for the observed distribution to be the outcome of the logically possible distribution of unique patterns over types. Rather, the strong

underrepresentation of type 5 patterns – only 1.6% (1 out of 64) of such expected patterns is attested – shows that the distribution is best explained in terms of the semantic and syntactic nature of the dependency relation, as captured by the scale in (22).

Tariana (Arawakan, Brazil) provides the sole counterexample: possessive modifiers, as shown in (23a), and verbal arguments, i.e. S_{AS} (23b), S_{PS} (23c), A_S and P_S (23d), do not trigger flagging on the head, while possessive and adpositional arguments are marked by an invariant prefix *i-*, as shown in (23e) and (23f). The pattern is limited to nominal dependents, and to a subset of relational nouns and postpositions.

- (23) a. *tfinu panisi*
 dog house
 'dog's house'
- b. *diha itfida di-eku-kha*
 ART turtle 3SG.NF-run-away
 'The turtle ran.'
- c. *diha ne:ri di-ñami-pidana*
 ART deer 3SG.NF-die-REM.PST.REP
 'The deer died.'
- d. *tfinu kuphe di-nitu-mahka*
 dog fish 3SG.NF-steal-REC.PST.NVIS
 'The dog stole the fish.'
- e. *ne:ri i-whida*
 deer INDF-head
 'the head of a deer'
- f. *ne:ri i-pumi*
 deer INDF-after
 'after the deer' (Aikhenvald 2003: 483, 667, 657, 296, 3, 3)

Note that this counterexample only targets head-marking; dependent-marking is fully in line with the hypothesis. Moreover, possessive modifiers and P_S are formally neutralized only via the absence of flagging: counterexamples making use of overt means of flagging are not attested in the sample.

Furthermore, it can be argued that the prefix *i-* is not a marker of the possessive dependency relation, but a mere dummy morpheme required to fill the prefix-slot of obligatorily (i.e. inalienably) possessed nouns in the absence of other morphology. In Tariana, pronominal possessive arguments are obligatorily indexed on the possessum head through prefixes; nominal possessive arguments

cannot be indexed, as a result of which the prefix *i-* has to be used as a semantically empty placeholder (the same analysis is argued for by Hengeveld & Mackenzie 2008: 412). Further support for this analysis comes from the fact that the prefix is also used in the absence of possessive arguments, i.e. it marks the non-possessed (and thereby non-indexed) form of obligatorily possessed nouns, e.g. *ñama i-sawi* (two INDF-horn) ‘two horns’ (Aikhenvald 1995: 183). A third argument in favor of this analysis is that the overt coding of inalienable possession but not of alienable possession, as (arguably) in Tariana, is cross-linguistically extremely rare (Haiman 1983, 1985: 130–136; Haspelmath 2006, 2008: 19–22; see also Chapters 2 and 3 of this thesis). Importantly, under a ‘dummy’ analysis of the prefix, the pattern illustrated in (23) does not constitute a counterexample.

Finally, it should be noted that the pattern in (23) is diachronically unstable, as the prefix is being lost under the influence of contact with Tucano languages that do without this prefix (Aikhenvald 2003: 124, 129–131). This results in the simple juxtaposition of possessor and possessum as in alienable possession, and occasionally in the use of indexes not only in the absence but also in the presence of nominal possessors. Notably, these innovative forms of possessive coding yield alignment patterns that are fully in line with the hypothesis.

Returning to the distribution of alignment types, it can be observed that patterns of type 1 (phrases vs. clauses, 56 instances) and type 3 (full neutralization, 54 instances) are most commonly attested. Type 3 alignment typically displays the absence of flagging; only Burushaski, Georgian, Tamil and West Greenlandic use a single (overt) means of flagging for dependents from each of the three categories. This can be attributed to the fact that some languages lack means of flagging altogether, while in many others dependency relations are overtly flagged only under certain referential or lexical conditions.

While the number of observed type 3 patterns does not deviate strongly from expectation (the factor difference is 1.15), type 1 patterns are attested roughly 6 times as often as expected, as shown in the final column of Table 14. Type 1&2 patterns display a similar asymmetry: they are (roughly) twice as frequent as expected. Type 2 patterns (modifiers vs. arguments), however, show the inverse kind of asymmetry: only 78% of all expected patterns are attested. Together, these data demonstrate that the sample languages strongly tend to formally neutralize dependency relations that are semantically distinct (i.e. the modifier/argument distinction, as in type 1 neutralizations) rather than those that are syntactically distinct (i.e. the phrase/clause distinction, as in type 2 neutralizations). Put differently: across the sample, the opposition between

phrasal and clausal dependency relations is expressed more often through flagging than the opposition between modifiers and arguments.

Interestingly, neutralizations of type 1 (phrases vs. clauses) minimally target possessive arguments: almost all type 1 and 1&2 patterns use the same flagging strategy – typically a single genitive case marker – for possessive modifiers and possessive arguments, or for all three phrasal dependents; the identical coding of possessive modifiers and adpositional arguments to the exclusion of possessive arguments is attested in only five type 1 and 1&2 patterns. In the sample as a whole, this type of formal neutralization is found in only nine alignment patterns of flagging. A likely explanation for the commonness of formal identity of possessive modifiers and possessive arguments is their shared possessor role. Diachronic factors may also be relevant: since adpositions often originate historically in relational nouns, such as body part terms, they are more likely to inherit the flagging of possessive arguments than that of possessive modifiers.

Finally, Table 15 below presents the distribution of the six alignment types over the 39 sample languages. It presents the alignment types (in the first column), the number of sample languages employing each type (in the second column), the number of alignment patterns attested per type (tokens, in the third column), and the number of unique alignment patterns – out of the 877 logically possible patterns – attested (types, in the final column). In the discussion of the table below, capitals are used for the alignment types in the first column and lower case is used for the unique alignment types in the final column. Seven sample languages are not part of the final column, as their alignment patterns could not be assigned to a unique type, due to the absence of adpositions or the indeterminacy of such a class.

The data show a highly homogeneous dispersion of alignment Types over sample languages, both in terms of tokens and in terms of types. In other words, the different alignment Types are found in many different languages; it is not the case that only a small number of languages are responsible for the frequent attestation of a certain alignment Type. This empirical observation is supported statistically, as there is a strong correlation between the distribution of sample languages and the distribution of alignment tokens ($r = 0.997$), and between the distribution of sample languages and the distribution of alignment types ($r =$

Table 15: The token, type and language distribution of alignment of flagging

Type	Frequency of languages	Frequency of attested alignment patterns (tokens)	Frequency of unique, attested alignment patterns (types)
1	24	56	10
1 & 2	17	35	10
2	12	27	11
3	23	54	16
4	1	2	2
5	1	1	1
Total	78	175	50

0.892).²⁵ Moreover, the distribution of alignment tokens correlates strongly with the distribution of alignment types ($r = 0.891$), showing that alignment Types that are cross-linguistically frequent, and thus have many tokens, strongly tend to belong to different unique patterns.

This section concludes the discussion of the results for the alignment of flagging. In the next section, I turn to the results for the alignment of indexing.

4.5.2 Alignment of indexing

Table 16 below presents the distribution of alignment types for indexing in the 39-language sample. Like Table 14 in the previous subsection, it presents the six alignment types (in the first column), the observed frequency of each alignment type (in the second column), the frequency of unique, logically possible patterns per type (in the third column), the expected frequency of each type given the actual number of observations (in the fourth column), and, finally, the factor difference between the observed and expected frequencies (in the final column).

²⁵ Correlations are computed using Pearson's correlation coefficient (r); the coefficient ranges from 0 to 1 (for positive correlations), with 0 indicating no correlation and 1 a total positive correlation (Johnson 2013: 302–307).

Table 16: The observed and expected distribution of alignment of indexing

Type	Observed frequency	Frequency of unique, logically possible alignment patterns	Expected frequency	Factor difference observed/expected frequencies
1	33	45	5.9	5.59
1 & 2	15	74	9.7	1.55
2	8	173	22.69	0.35
3	57	235	30.82	1.85
4	1	30	3.93	0.26
5	1	320	41.96	0.02
Total	115	877	115	-

The table demonstrates that indexing alignment of type 5 is cross-linguistically rare: only one such pattern is attested, while there are 320 logically possible patterns of this type. Alignment of type 4 is also attested once, but this is unsurprising given the low number of logically possible patterns of this type (only 30 out of 877, i.e. 3%). Other alignment types are either much more frequent (type 1, with 33 instances, and type 3, with 57 instances) or are not as strongly underrepresented as type 5 (this concerns type 2, which has a factor difference of 0.35). This distribution of alignment types confirms the main hypothesis for indexing: if modifiers in phrases are indexed in the same way as arguments in clauses, arguments in phrases are expressed by this indexing strategy as well. The use of a single indexing strategy for dependency relations that are both semantically and syntactically distinct, as in type 5 patterns, is strongly dispreferred across the languages of the sample. Hence, the data show that indexing strategies, i.e. individual indexes or their absence, always cover continuous segments of the scale repeated in (24):

(24) modifier in phrase – argument in phrase – argument in (main) clause

This empirical finding is supported by a chi-square goodness-of-fit test, which demonstrates that the difference between the observed distribution of alignment types (in the second column) and the expected distribution of alignment types

(in the fourth column) is statistically significant ($\chi^2 = 201.29$, $p < 0.0001$).²⁶ This finding demonstrates that the modifier/argument opposition and the phrase/clause opposition both play an important role in the cross-linguistic encoding of dependency relations, not only in terms of flagging, but also in terms of indexing.

Interestingly, the sole counterexample for indexing in the data stems from the same language as the sole counterexample for flagging: Tariana. As shown below, possessor modifiers (25a) and P arguments (25b) are never indexed in Tariana, as opposed to possessor arguments (25c), postpositional arguments (25d), S_{AS} (25e), S_{PS} (25f) and A_s (25b). This type 5 pattern only applies to pronominal dependents, and to a subset of relational nouns and postpositions.

- (25) a. *naha amaku-pe*
 3PL hammock-PL
 ‘their hammocks’
- b. *na-na na-pinita-ka-ta*
 3PL-OBJ 3PL-follow-DECL-REPV
 ‘They were following them.’
- c. *na-whe-ri-nipe*
 3PL-grandparent-M-PL
 ‘their grandfathers’
- d. *na-dalipa*
 3PL-near
 ‘near them’
- e. *na-musu na:*
 3PL-go.out 3PL.go
 ‘They started going out.’
- f. *na-yami-na*
 3PL-die-REM.PST.VIS
 ‘They died.’ (Aikhenvald 2003: 652, 470, 634–635, 533, 438, 413)

Like the type 5 pattern of flagging in Tariana, this pattern is based on the absence of formal coding; no sample language uses a single overt index-set for phrase modifiers and clause arguments, but not for phrase arguments.

²⁶ Expected frequencies are computed in the same way as explained in the previous subsection, e.g. 45 of out 877 possible indexing patterns belong to type 1, which, given a total sample size of 115 indexing patterns, yields an expected number of 5.9 patterns of this type.

Focusing on the distribution of alignment types 1 to 4, it can be observed that indexing alignment of type 3 is most common (57 instances) and is almost twice as common as expected (cf. the final column of Table 17). It happens almost exclusively on the basis of the absence of indexing (in 51 out of 57 patterns, i.e. 90%); only Abkhaz, Macushi, Nivkh and Turkish use overt means of indexing for each of the three types of dependency relations. As with flagging, indexing patterns of type 1 show the largest difference between observed and expected frequencies: this type is more than 5 times as common as expected. Type 1&2 patterns are also more frequent than expected, though only slightly: 1.5 times. The inverse asymmetry applies to patterns of type 2: only 3.5% of all expected patterns are attested in the data. In sum, these asymmetries demonstrate a clear preference for the formal neutralization of the modifier/argument opposition (in type 1 neutralizations) as opposed to the phrase/clause opposition (in type 2 neutralizations). Hence, the opposition between phrasal and clausal dependency relations is expressed more often than the opposition between modifiers and arguments, in flagging (as shown in the previous subsection) as well as in indexing. Moreover, as with flagging, type 1 neutralizations where possessor modifiers and adpositional arguments are expressed with the same indexing strategy to the exclusion of possessor arguments are rare: only three such type 1 and 1&2 patterns are attested. The same synchronic motivation – the shared role of possessor modifiers and possessor arguments – and diachronic motivation – the shared historical origin of possessor arguments and adpositional arguments – given for flagging may apply here.

Finally, Table 17 presents the distribution of alignment types (in the first column) in terms of the number of sample languages (in the second column), the number of attested alignment patterns (tokens, in the third column) and the number of unique, attested alignment patterns (types, in the final column). As with Table 15, capitals are used for alignment types in the first column, while lower case is used for alignment types in the final column. The seven sample languages excluded from Table 15 are also excluded here.

The data show strong correlations between the individual columns: the distribution of sample languages correlates strongly with the distribution of tokens ($r = 0.995$) and with the distribution of types ($r = 0.976$). Hence, cross-linguistically common alignment Types, both in terms of alignment patterns in general (tokens) and in terms of unique alignment patterns (types), are widely dispersed among the different sample languages. Also note that token frequency correlates strongly with type frequency ($r = 0.964$), showing that alignment

Types with a high number of tokens also belong to a high number of unique types.

Table 17: The token, type and language distribution of alignment of indexing

Type	Frequency of languages	Frequency of attested alignment patterns (tokens)	Frequency of unique, attested alignment patterns (types)
1	16	33	8
1 & 2	10	15	6
2	5	8	6
3	31	57	16
4	1	1	1
5	1	1	1
Total	67	115	41

In sum, this section has shown that alignment patterns are sensitive to the semantic nature of dependency relations (head-modifier relations vs. head-argument relations) as well as to their syntactic type (phrasal vs. clausal). Specifically, it is shown that if (possessor) modifiers in phrases receive the same flagging or indexing strategy as arguments in clauses, (possessor or adpositional) arguments in phrases receive this coding strategy as well. In the next section, I show that adpositional phrases play a similar bridging role to phrase arguments in general, but then in the formal neutralization of arguments in possessive NPs and arguments in main clauses.

4.6 The bridging role of adpositional phrases

In the previous section it was demonstrated that arguments in phrases play a bridging role in the formal neutralization of modifiers in phrases and arguments in clauses. Interestingly, adpositional phrases play the same bridging role, but then among arguments only. Specifically, the data point to the following generalization: if arguments in possessive NPs are expressed in the same way as (one or more) arguments in main clauses, arguments in adpositional phrases receive this formal treatment as well. This generalization applies to both flagging and indexing, discussed in turn below.

Table 18 provides the relevant data for flagging. The first and second columns present the five distinct ways in which possessor arguments (PSR ARG), adpositional arguments (ADP ARG) and clausal arguments (CLS ARG)

can be formally neutralized, which are labeled types A to E; brackets separate formally identical dependency relations. The table also presents the observed distribution of each type (in the third column), the distribution of the 877 unique, logically possible alignment patterns over the five types (in the fourth column), the expected frequency of each type (in the fifth column) and finally the factor difference between the observed and expected frequencies (in the final column).²⁷

Table 18: The observed and expected distribution of flagging alignment of arguments

Type	Patterns of identical coding	Obs. freq.	Freq. of unique, possible alignment patterns	Exp. freq.	Factor difference obs./ exp. freq.
A	(POSS ARG, ADP ARG) (CLS ARG)	29	67	12.07	2.40
B	(POSS ARG) (ADP ARG, CLS ARG)	35	173	31.17	1.12
C	(POSS ARG, ADP ARG, CLS ARG)	45	136	24.5	1.84
D	(POSS ARG) (ADP ARG) (CLS ARG)	24	83	14.95	1.60
E	(POSS ARG, CLS ARG) (ADP ARG)	25	418	75.31	0.33
Total		158	877	158	-

Comparing the observed frequencies with the expected ones shows that all alignment types are attested more often than expected, except for one: type E, which is strongly underrepresented. As shown in the final column, only 33% of all expected type E patterns are attested. This asymmetry demonstrates that the identical flagging of possessor arguments and (one or more) clausal arguments to the exclusion of adpositional arguments is dispreferred cross-linguistically. A chi-square goodness-of-fit test shows that the difference between the observed and the expected frequencies is statistically significant, supporting this empirical finding ($\chi^2 = 80.44$, $p < 0.0001$). On the one hand, this finding can be explained by the phrasal vs. clausal nature of the relevant dependency relations: possessor arguments, as parts of phrases, can only be encoded as arguments in clauses, when other arguments in phrases, i.e. adpositional arguments, receive this

²⁷ The expected frequency of each type is computed in the same way as in the previous section, but note that the number of unique, possible patterns per type differs, since a different set of alignment patterns is being investigated here. As in testing the main hypothesis, formal neutralizations need not affect all clausal arguments. Patterns in which possessor arguments receive the same coding as (one of more) clausal arguments are classified as the unpredicted type E, irrespective of additional formal identities. This maximizes the *a priori* likelihood of finding a counterexample.

coding strategy as well. On the other hand, diachronic factors may be relevant: since adpositions often historically derive from body part nouns and other relational nouns, they are likely to receive the same coding as possessor arguments; however, adpositions also commonly derive from verbs in serial verb constructions, which may motivate their formal identity with clausal arguments. Generalizations regarding the type of coding used, discussed below, support these motivations. In general, this finding demonstrates the typological relevance of the phrasal vs. clausal nature of dependency relations, not only across modifiers and arguments (as demonstrated in the previous section), but within the domain of arguments as well.

Focusing on the type of flagging strategies used, it can be observed that adpositional arguments are often formally identical to possessor arguments and/or clausal arguments (as in types A, B and C) due to the absence of flagging. By contrast, the formal identity of adpositional arguments and possessor modifiers, or possessor modifiers **and** possessor arguments is typically revealed through overt means of flagging, such as a shared genitive case marker. This tendency can be explained by the fact that possessor arguments are often unmarked, especially in languages with an alienability split, which in turn follows from the fact that they are semantically inherent to the head noun and therefore need no overt coding of their possessor role (cf. Chapters 2 and 3 of this thesis). Again, diachronic factors may apply: adpositions may inherit the (absence of) flagging of possessive arguments – given their historical origin as relational nouns – or from verbs in serial verb constructions where the clausal argument typically remains unmarked. Note, however, that adpositional and clausal arguments may also receive the same overt means of flagging; interestingly, this typically involves the P, e.g. a single accusative or oblique case marker, rather than the A, e.g. a single ergative case marker. This is consistent with findings in other types of typological research, such as the relationship between adpositional phrases and P arguments in terms of word order (Dryer 2013). A relevant example is given in (26) for Nama (Khoe-Kwadi, Namibia):

- (26) a. *'oms-à xuú*
 house-ACC from
 'from the house'
- b. *'áo-p ke 'àrí-p-à kè mũũ*
 man-3M.SG DECL dog-3M.SG-ACC PST see
 'The man saw the dog.' (Hagman 1977: 193, 148)

Finally, note that the unique flagging of adpositional arguments (as in alignment

of types D and E) is typically characterized by overt flagging: zero-flagging is attested in only 3 of the 49 patterns.

The bridging role of adpositions also surfaces in indexing; Table 19 presents the relevant data, in the same manner as Table 18.

Table 19: The observed and expected distribution of indexing alignment of arguments

Type	Patterns of identical coding	Obs. freq.	Freq. of unique, possible alignment patterns	Exp. freq.	Factor difference obs./ exp. freq.
A	(POSS ARG, ADP ARG) (CLS ARG)	15	67	7.18	2.09
B	(POSS ARG) (ADP ARG, CLS ARG)	18	173	18.54	0.97
C	(POSS ARG, ADP ARG, CLS ARG)	48	136	14.58	3.29
D	(POSS ARG) (ADP ARG) (CLS ARG)	4	83	8.9	0.45
E	(POSS ARG, CLS ARG) (ADP ARG)	9	418	44.80	0.20
Total		94	877	94	-

The table demonstrates that type E is again the most underrepresented type: only 20% of the expected patterns of this type are attested in the data. Type D patterns are also underrepresented, but less strongly so: 45% of the expected patterns of this type are attested. A chi-square goodness-of-fit test demonstrates a significant difference between the observed and the expected distribution of types ($\chi^2 = 116.47$, $p < 0.0001$).

Importantly, type E patterns are significantly underrepresented in the sample as a whole, while type D patterns are not. This can be demonstrated by two Fisher's Exact tests. The first compares the observed and expected frequencies of type E patterns with the observed and expected frequencies of all other alignment types, as represented by the following 2x2 contingency table:

Table 20: The observed and expected frequency of type E vs. other alignment types

	Observed frequency	Expected frequency
Type E	9	45
Types A, B, C, D	85	49

The p value for this table is < 0.0001 (one-tailed), showing that type E patterns are indeed significantly underrepresented compared to the other alignment types. The second Fisher's Exact test shows that this does not apply to alignment patterns of type D, for which a similar 2x2 contingency table can be given:

Table 21: The observed and expected frequency of type D vs. other alignment types

	Observed frequency	Expected frequency
Type E	4	9
Types A, B, C, D	90	85

This table yields a *p* value of 0.13 (one-tailed), showing that type D patterns are not significantly underrepresented compared to the other alignment types attested. In sum, this demonstrates that type E patterns are indeed cross-linguistically rare, and thus that languages tend to avoid using the same indexing strategy for possessor arguments and clausal arguments without using this indexing strategy for adpositional arguments as well. This in turn supports the bridging role of adpositions in the formal neutralization of possessor arguments and clausal arguments across the languages of the sample.

Turning to the type of indexing strategies used, the data show that the formal identity of adpositional arguments and possessor arguments (as in type A and C patterns) is often achieved through the absence of indexing. This could be due to the fact that the indexing of adpositions is relatively uncommon cross-linguistically: it is found in 28% of the languages in Siewierska (2004) and in 34% (11 out of 32) of the languages with adpositions in my sample. However, the use of identical indexes for these two arguments is also attested, which can be attributed to the fact that possessor arguments are frequently indexed as opposed to possessor modifiers (cf. Siewierska 2004: 143-145; Stassen 1997: 38). The frequent indexing of possessor arguments may in turn be related to the argumenthood of such possessors via grammaticalization: as semantically obligatory dependents of the head, possessor arguments are arguably expressed more often than possessor modifiers, and therefore more likely to grammaticalize into indexes on the head (cf. Nichols 1992: 121–122; Haspelmath 2008: 19–22). Adpositional and verbal arguments also often remain unindexed (as in type B and C patterns), but may share the same overt indexing strategy as well. Interestingly, indexes on adpositions strongly tend to be identical to those used for P arguments, sometimes also used for S_P arguments, rather than A or S_A arguments. A possible explanation lies in the semantic connection between patientive arguments and possessor arguments, being items that have no control over the relationship to their head, which in turn explains their common identical expression across the world's languages (cf. Seiler 1983a: 22, 1983b; Wilson 1982: 123–130). Given that relational nouns are common diachronic sources of adpositions, this may explain why adpositional and clausal arguments often share the same overt indexing strategy.

In sum, this section has shown that possessor arguments and clausal arguments are rarely coded alike without the additional inclusion of adpositional arguments. Adpositional arguments thus play a bridging role in the formal neutralization of possessor arguments and clausal arguments, which can be explained by the shared phrasal nature of possessive NPs and adpositional phrases on the one hand, and the diachronic link between adpositions and possessum nouns, and between adpositions and verbs on the other. This finding shows that the phrase/clause distinction is relevant, not only in the coding of dependency relations in general (as shown in the previous section) but also in the coding of arguments in particular.

4.7 Conclusions

Alignment is traditionally studied as the morphosyntactic coding of verbal arguments in main clauses. In this paper, I study alignment as a notion that transcends the domain of the clause, as motivated by a semantic distinction between modifier dependents and argument dependents, which applies to both phrases and clauses. The aim of this paper is to determine to what extent the alignment of phrasal and clausal dependency relations is sensitive to the opposition between modifiers and arguments on the one hand, and the phrasal vs. clausal nature of the dependency relation on the other. To this end, I studied the morphosyntactic coding of possessive NPs, adpositional phrases and the core arguments of one- and two-participant verbs in main clauses, in a worldwide sample of 39 languages. In so doing, I took into account language-internal variation in alignment conditioned by referential and (partly) lexical factors, following recent work on clausal alignment typology. I hypothesized that languages show significant variation in alignment patterns, but that one pattern is cross-linguistically rare: the identical coding of modifiers in phrases and arguments in clauses without the inclusion of arguments in phrases. This hypothesis is fully borne out by the data, providing evidence for the following cross-linguistic tendency:

- (i) If modifiers in phrases are encoded in the same way as arguments in clauses, arguments in phrases receive the same formal treatment.

This generalization shows that arguments in phrases play a bridging role in the formal neutralization of modifiers in phrases and arguments in clauses, which is motivated by the phrasal nature of the dependency relation that they share with modifiers in phrases, and their argumenthood shared with arguments in clauses.

Additionally, it was shown that adpositional phrases play a similar bridging role, but then among arguments only. The data yield the following cross-linguistic tendency:

- (ii) If possessor arguments (in phrases) are coded in the same way as arguments in clauses, adpositional arguments (in phrases) receive the same formal treatment.

This finding is motivated by the shared phrasal nature of dependency relations containing possessor arguments and adpositional arguments, and by the diachronic link between adpositions and (relational) possessum nouns, on the one hand, and between adpositions and verbs in serial verb constructions, on the other.

To conclude, this study has shown that alignment is a phenomenon pertaining not only to clauses, but to phrases as well. Specifically, it has been shown that alignment patterns are sensitive to the semantic nature of the dependency relation (head-modifier relations vs. head-argument relations) as well as to the syntactic nature of the dependency relation (phrasal vs. clausal). Both the modifier/argument opposition and the phrase/clause opposition have been shown to be two relevant typological parameters in the morphosyntactic coding of dependency relations.

In order to further corroborate these findings, future research should expand the sample of languages and the constituent types investigated in this study. Ideally, such research will also shed more light on the cross-linguistic relevance of referential properties in possessive and adpositional coding, which remain underresearched, and their relationship to referential conditions on clausal argument coding. Finally, future research should look beyond the prototypical noun and verb classes considered in this study, as this may yield a richer typology of alignment patterns to be observed in the world's languages.