

Chapter 28

LFG and Austronesian languages

I Wayan Arka

The Australian National University

Li-Chen Yeh

The Australian National University

Austronesian (AN) languages are known for their diverse grammatical characteristics in many typological and descriptive works. Their properties provide fertile grounds for testing assumptions in syntactic theories. In this chapter, we demonstrate that the parallel correspondence architecture of LFG can be used as a powerful tool for language-specific linguistic analysis, while also precisely capturing the cross-linguistic differences within and between Western and Eastern AN languages. LFG is flexible in incorporating analytical tests, such as adverbial insertion and clitic placement for examining constituency; reflexive binding, nominal marking and pronominal-indexing for syntactic status of an argument. Although AN languages have posed challenges to traditional syntactic notions of subject, as well as the mapping between grammatical relations and functions, we show that such multi-dimensional views of grammar, and projection design, can deal with these challenges efficiently, and also lead to a coherent comparative representation of AN languages for the purpose of tracking morphosyntactic stages according to their respective typological categories.

1 Introduction

The world of Austronesian (AN) languages comprises a huge and diverse language family, which covers a wide geographical span ranging from Formosan languages in the northwest of the Pacific, Malagasy at its westernmost point, Māori in south Oceania, Hawaiian in the northeast and Rapanui at its most eastern point. In fact, this geographical spread is a historical outcome of the prehistoric settlement by AN speaking communities (Pawley & Pawley 1998; Bellwood



2007: 242). The histories of the people in the widespread Asian-Pacific region are testimony to the genealogical continuity of AN languages that form one single language family.

Initially, the AN dispersal began from the island of Taiwan in the northern part of the Pacific island chain, outside the east Asian continental mainland (Pawley 2002, Bellwood 2007, Skoglund et al. 2016, Blust 2019). The languages natively spoken on the island of Taiwan are direct descendants of the Proto-AN language (Blust 1999). These languages are collectively called FORMOSAN LANGUAGES, and are sisters to the common ancestor of the remaining AN languages, Proto-Malayo-Polynesian (PMP).

The AN expansion took place in subsequent waves, as laid out chronologically below (Adelaar 1989, Bellwood 2007: 201-254). The PMP subgroup began to split up as it spread from Taiwan to the Philippines (est. 2200 BCE), along with early migration settling in Micronesia. There were also dispersals from the Philippines into Indo-Malaysia and eastward to New Guinea and the Bismarck Archipelago. The settlement in the Bismarck has been dated to around 1350 BCE. Later on, the dispersal went further eastwards into the Pacific (e.g. Solomons, Vanuatu, New Caledonia, Fiji, Tonga and Samoa) between 1200 and 900 BCE. After 600 CE, the AN occupation of eastern Polynesia occurred, and immigrants from Borneo arrived in Madagascar around 500 CE.

The AN migration history above shows the geographical distribution of AN languages. Meanwhile, it similarly indicates a diachronic progression in the varying prototypical features of AN morphosyntax and offers a general guideline for the typology of AN languages. A major typological distinction can be made between WESTERN AN and EASTERN AN.

The Western AN group¹ includes languages of Taiwan, the Philippines, western Indonesia, Malaysia and Madagascar. These languages are typically characterised by their robust ‘alternating’ and ‘symmetrical voice’ systems, which use verbal morphology to mark a non-Agent argument as grammatical SUBJ(ect) or PIVOT without demoting the Agent argument to oblique (discussed in Section 4). This non-demotion property of the Western AN voice system differs from the commonly observed active-passive voice alternation system in Indo-European languages like English.

The symmetrical voice systems in Western AN pose challenges to many grammatical frameworks, including some versions of the standard Lexical Mapping

¹In this chapter, Western AN is used as a cover term for symmetrical voice languages. It differs from Himmelmann’s (2005) geographic label of ‘Western Austronesian’ which encompasses all non-Oceanic languages.

Theory analysis in LFG. The non-demotion property of the symmetrical voice system licenses a passive-like structure with a non-actor thematic role selected as subject, called Patient (or Undergoer) voice, but unlike in the passive, the non-subject actor role has the most prominent core status. This leads to a mismatch between its semantic and syntactic prominence: the most prominent semantic (agent) argument is not the most privileged (SUBJ) argument. The diagnostic tool for identifying this mismatch involves reflexive binding (see Section 2). The surface realisation of reflexive binding allows reflexive pronouns to be bound by antecedents bearing both the least and most prominent grammatical functions. While posing challenges to many grammatical frameworks, this unusual and intricate variation of voice alternation is best explained by the multi-layered argument structure of LFG's architecture which tackles associations between grammatical functions and grammatical relations.

The Eastern AN language group² includes languages of Timor-Leste, New Guinea and Oceania. In contrast to symmetrical-voice languages, Eastern AN languages no longer maintain the layered distinctions at the semantics-morphology-syntax interface. We refer to these languages as the non-alternating type because the typical alternative selection of a semantic role as grammatical SUBJ, as seen in Western AN languages, is not observed. Instead, Eastern AN languages are characterised by other properties such as the emergence of systematic pronominal indexing, as well as increased complexity in other parts of its grammar as seen in their rich serial verb constructions (SVCs) and clausal complementation (see Section 5). The pronominal indexing paradigm is an exclusive feature of Eastern AN languages, and in this regard, they may be referred to as indexing-type languages.

Indexing-type languages show distinct properties in complex constructions that are intriguing for typological comparisons and important for theoretical testing. These languages show a striking consistency in the distinction of (x)COMP (i.e. clausal complementation with(out) shared missing SUBJ: see Vincent 2023 [this volume] and Section 5 of this chapter) while also revealing a significant difference with regard to the structural tightness between regular complementation and complex predicates. The latter has usually been subsumed under the general heading of SVCs. It is not always straightforward in many syntactic theories to capture the distinction among different kinds of (x)COMP (e.g. control, SVCs and multi-verb constructions in coordination and subordination). Nonetheless, a few clear cases of the distinction in argument gapping strategies can be effectively demonstrated from the LFG perspective where, crucially, no movement is

²Eastern AN languages have been commonly referred to in the literature as preposed-possessor languages, Oceanic languages or isolating languages of eastern Indonesia.

required (cf. Sells's (2023 [this volume]) comparison of LFG with the traditional analysis of control/raising complementation in transformational grammars). Instead, the LFG perspective can clearly present how a verb form with/without an overt voice marker can serve as a diagnostic tool for testing the core status of (x)COMP, and how voice morphology (as well as negation) forms a criterion for teasing apart the differences between complementation and SVCs. Even the compound-like structure of the complex event-composition in SVCs can be captured via the interrelated specifications on different linguistic dimensions (see Section 5.2).

Even though the above description has provided a general indication of the major typological differences between Western and Eastern AN languages, two important points should be made on the typological diversity of AN morphosyntax.

First, not all Western AN languages behave alike. Symmetrical-voice languages are typically further subcategorised into 'Philippine-type' and 'Indonesian-type', due to their distinct characteristics in word order (cf. Section 3), the number of semantic roles allowed as privileged arguments (cf. Section 4.2), and the use of case-marking flagging and applicative constructions (cf. Section 4.3). Although most Western AN languages may be subcategorised further, some transitional languages do not adhere to the typological profile of either type (Kroeger 2023). Certain Western AN languages in Taiwan even appear to show disputable traits of asymmetry in syntax (cf. Section 4.4).

Moreover, geography and typology do not always neatly align. For instance, certain Barrier Island languages off the south coast of Sumatra, such as Mentawai (Lenggang et al. 1978, Arka 2006), Enggano (Crowley n.d., Hemmings in preparation) and Nias (Brown 2001), do not show a symmetrical voice property of the type seen in the Western AN group, but they have developed person-marking prefixes on verbs that encode subject similar to NOM(inative) subject prefixes in outlier AN languages in southern/eastern Indonesia, such as Kambara in Sumba (Klamer 1998) and Woi in West Papua (Sawaki 2016). Makassarese, spoken in Sulawesi, has unmarked word order like the Philippine-type, but it also exhibits systematic pronominal indexing on the verbal predicate. In Makassarese, a transitioning state of word order change is observed in the expression of contrastive focus through clefting (cf. Section 6).

Typologically, the AN language family is intensely diverse with a variety of transitioning languages comprising two heterogeneous macrogroups. This diversity has posed difficulties for descriptive and comparative analysis, particularly for long-standing and often controversial topics of typological and theoretical significance, such as ergativity or the complex interconnection between sur-

face grammatical relations and deeper semantic-syntactic argument structures. Nonetheless, LFG's parallel correspondence architecture provides the necessary flexibility for a coherent and comparable descriptive representation of AN grammar on these topics.

In this chapter, we show how LFG can be used as a descriptive and analytical tool to capture the typological range of AN linguistic diversity on these selected topics. We begin by highlighting the LFG modular design, and its application in modelling the morphosyntactic operation of AN voice systems (Section 2), then illustrate how the LFG framework can capture word order variation (Section 3), grammatical functions and alignment (Section 4), complex argument sharing constructions (Section 5) and information structure (Section 6).

2 LFG modular design and Austronesian linguistics

LFG is modular in its design. From the LFG perspective, a language is construed as multiple dimensions of linguistic information, and each dimension constitutes an individual module, or structure, that comes with its own formal properties. Different structures are parallel but are linked by principles of correspondence, as introduced in Belyaev 2023b [this volume].

In the standard LFG framework, traditional syntactic structure is primarily represented on two structural levels: CONSTITUENT STRUCTURE (c-structure) and FUNCTIONAL STRUCTURE (f-structure). Ordering of constituents and syntactic categories are analysed in c-structure, whereas grammatical functions (GFs) of arguments and grammatical features are dealt with in f-structure, as detailed in Belyaev 2023a [this volume]. In subsequent developments, semantic ARGUMENT STRUCTURE (a-structure) was proposed in Lexical Mapping Theory (LMT) to capture cross-linguistic GF alternations (Bresnan & Kanerva 1989); see also Findlay et al. 2023 [this volume]. Bresnan & Kanerva (1989) propose that a-structure is represented as a list of semantic roles which are directly mapped onto GFs. Bresnan & Kanerva's (1989) LMT works well to account for voice and alternative argument realisations in languages like English; e.g. the agent's SUBJ(ect)-OBL(ique) alternation in passivisation.

However, the rich voice systems of western AN languages pose a problem for this version of LMT such that semantic a-structure and the traditional analysis of GFs cannot be maintained. Based on data from western AN languages, which will be discussed in detail in this chapter, we argue against Bresnan & Kanerva's (1989) version of mapping theory; see Arka (2003a: 119–124) for a comprehensive examination of the evidence and justification. Consider the following examples from Balinese (1) and Puyuma (2):

(1) Balinese (WMP, Indonesia) (Arka, own knowledge)³

- a. [Tiang]_{SUBJ} ng-ejang [nasi]_{OBJ} [*(di) bodag-e]_{OBL_{LOC}}.
 1SG AV-put rice in basket-DEF
 'I put rice in the basket.'
- b. [Nasi-ne]_{SUBJ} Ø-ejang [tiang]_{OBJ} [*(di) bodag-e]_{OBL_{LOC}}.
 rice-DEF UV-put.rice 1SG in basket-DEF
 'I put rice in the basket.'

(2) Puyuma (Formosan) (Teng 2008: 47-48)

- a. Trakaw [dra paisu]_{OBL} [i isaw]_{SUBJ}.
 <AV>steal INDF.OBL money SG.NOM Isaw
 'Isaw stole money.'
- b. [Tu=]_{OBJ} trakaw-anay [i tinataw]_{SUBJ} [dra paisu]_{OBL}.
 3GEN= steal-CV SG.NOM his.mother INDF.OBL money
 'He stole money for his mother.'

The examples above represent two salient features of the AN voice system and related argument realisations. First, verbal voice morphology marks SUBJ selection (cf. Section 4). The ACTOR VOICE (AV), indicated by *ng-* in Balinese (1a) and <*em*> in Puyuma (2a), selects the most agent-like role, or A (*tiang* and *Isaw* respectively) as SUBJ.⁴ The UNDERGOER VOICE (UV) in Balinese is indicated by a zero prefix and selects a patient-like (P) role as SUBJ as in (1b), whereas the CONVEYANCE VOICE (cv) in Puyuma selects a peripheral role, as in the beneficiary 'mother' in (2b)⁵ (see Section 4.2 for the properties of SUBJ/PIVOT and the typology of AN voice

³In this chapter, a language is presented with its linguistic and geographical classification in the first instance based on information from Glottolog for the purposes of locating genealogical and typological relations between languages. WMP and CEMP stand for Western Malayo-Polynesian and Central-Eastern Malayo-Polynesian respectively.

⁴Following standard conventions in language typology (Comrie 1978, Dixon 1979, Croft 2003, Haspelmath 2007, among others), we adopt the following abbreviations to denote generalised semantic roles: A represents the argument that is most actor-like, while P represents the argument that is most patient-like in a transitive predicate. It is worth noting that P is approximately synonymous with the undergoer (U) macro-role in Role and Reference Grammar (Foley & Van Valin 1984).

⁵In AN linguistics, non-AV or UV is also often called Objective Voice (Kroeger 1993). It is the voice type that selects certain semantic roles other than the actor as SUBJ/PIVOT. In the AN languages of the Philippines and Taiwan, there are typically different types of UV named after the associated semantic role of the SUBJ and each has its own verbal morphology, e.g. *-anay* for Conveyance Voice (cv) in Puyuma in (2).

systems). The voices in these AN languages are called SYMMETRICAL VOICE. Evidence for their symmetry comes from the fact that agent and non-agent roles are equally selectable by the voice morphology as the “privileged argument”, which is analysed as SUBJ/PIVOT in LFG (Kroeger 1993, Manning 1994, Arka 2003a, Falk 2006). In addition, symmetry is seen in verbal marking, particularly in Puyuma, where different voice types (e.g. AV and CV) are equally marked.⁶

Second, non-AV alternations in (1b) and (2b) are not passivisation. That is, in both Balinese UV and Puyuma CV, the promotion or selection of a non-A role as SUBJ is not accompanied by the demotion of A to OBL, a lower ranked function. Functionally, the A argument is the OBJ(ect); it retains its core status in the structure. This is clearly seen in Balinese, where the A of the UV structure in (1b) appears immediately after the verb as a bare Noun Phrase (NP), like the OBJ of the AV verb in (1a). Note that OBL in Balinese is flagged by a preposition. In addition, Balinese does have a passive, in which case the agent appears as OBL (see Arka 2003a). Likewise, the A of the CV verb in Puyuma is realised as a bare GEN(itive) clitic. A free OBL nominal in Puyuma is also prepositionally flagged, and if it is a pronominal, it has a special OBL form distinct from the GEN or NOM(inative)⁷ form (see Teng 2008: 63). Bresnan & Kanerva’s (1989) classic LMT approach cannot account for the agent’s alternative realisation as OBJ in UV, as in Balinese (1b) and CV in Puyuma (2b), since the agent is inherently classified as [–o] (i.e. not object-like), thus only allowing for the SUBJ-OBL alternation as seen in passives.

To capture the non-demotion of A in UV and other salient typological and morphosyntactic properties of AN voice alternation in LFG, we distinguish GFS from GRS (Grammatical Relations). GRS are clause-internal relations that reflect semantic-syntactic dependency between a predicate and its dependents. They form the so-called syntacticised a-structure in Manning (1994), Arka (2003a) and Arka & Manning (2008). This syntacticized a-structure, as distinct from the semantic a-structure in Bresnan & Kanerva’s (1989) classic LMT, incorporates syntactic information regarding coreness/obliqueness alongside its structural prominence, which includes thematic ranking. Although GRS and GFS belong to different layers of structure, the two are interrelated and are mapped together through

⁶Balinese and Puyuma belong to two different subcategories of Western AN languages: the Indonesian-type and Philippine-type, respectively. They differ in their number and type of voice distinctions, and the syntactic properties of their non-SUBJ arguments (e.g. the obliqueness of P in AV) (cf. Section 4). Note that the UV verb in Balinese is also analysed as being ‘marked’; it is realized as a zero UV prefix on the basis of its contrasting form with the AV verb.

⁷Note that Puyuma is an ergative language. In Teng (2008), NOM refers to the case assigned to SUBJ. It should not be confused with nominative case in NOM-ACC languages. See the discussion in Section 4.4.

linking principles. In AN languages, the mapping is marked by verbal voice marking and/or indexing, depending on the available coding resources. Although the GF-GR distinction may not be assumed and embraced in all LFG analyses, we contend that it is essential to provide a consistent explanation for voice alternations and related alternative argument realizations in languages that exhibit both accusative and ergative properties, as seen in western Austronesian languages such as Balinese and Indonesian.

We use different convention labels for GRS and GFs to avoid confusion: small caps for GFs (e.g. SUBJ, OBJ) and lower case for GRS (e.g. subject, object). We illuminate the significance of the GF-GR distinction by taking into account recent findings (e.g. reflexive binding) in descriptive and typological research in AN linguistics. The informal GR or a-structure representation is given in (3a) for the Balinese verb *jang* ‘put’, and its simplified representation is given in (3b). The a-structure in (3) outlines that *jang* is a three-place predicate with core arguments (i.e. agent and patient) and a third non-core oblique locative (LOC) (i.e. location). In our analysis, we adopt the convention of using a vertical line (|) to explicitly delineate argument classes, distinguishing between core and oblique or non-core arguments within the argument structure list. This convention supplements the valency information specified in the a-structure representation in the lexicon.⁸

(3) Syntacticised argument structure of *jang* (Balinese):

- a. 'put< ARG1:agent , ARG2:patient | ARG3:location >'
(subject) (object) (oblique)
- b. 'put< 1:agt, 2:pt | 3:loc>'

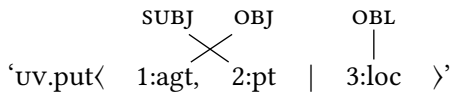
The following should be noted regarding the representation of GRS in (3). First, GRS conceptually reflect event construal and participant roles, which signify ‘who

⁸Syntactic coreness/obliqueness and valency are distinct syntactic properties that pertain to whether an argument is core or not and the number of arguments a predicate takes, respectively. These properties are not always predictable from the semantics of the predicate, as evidenced by the variation observed across languages when expressing the same event using equivalent verbs with different syntactic argument structures. For example, the verb *give* is semantically a three-place predicate involving a giver, givee, and gift. However, languages differ in how they syntactically realize these arguments. Balinese *baang* 'give' exclusively permits ditransitive constructions without a dative alternation (Arka 2003b: 63–63), while English allows the verb 'give' to function as either ditransitive, like in Balinese, or monotransitive with an oblique as the third argument, depending on the context. Furthermore, languages such as Indonesian exhibit ditransitive/transitive alternations facilitated by applicative morphology. Consequently, this lexically specific syntactic information must be encoded in the syntacticized *a*-structure rather than the semantic structure within the lexicon. Arka's (2003a) work provides detailed evidence from Balinese supporting this perspective.

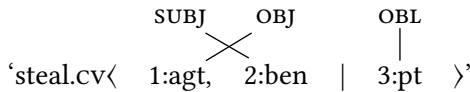
does what to whom'. This is the semantic-conceptual basis underlying valency and transitivity information in the syntacticised a-structure. The valence information specifies the number of arguments (e.g. one, two, or three arguments) and syntactic/semantic transitivity specifies types of arguments (i.e. core or oblique, and the associated semantic roles). The Balinese verb *jang* in (3) is a three-place transitive predicate with two core arguments, 1:agent and 2:patient, and one non-core argument, 3:location. For the purposes of comparative typology, 1:agent and 2:patient will be referred to as *subject* and *object*, respectively (noting lower case). They roughly correspond to the typologists' labels A and P/O respectively), which are distinguished from surface GFS, SUBJ and OBJ.

Voice morphology on the verb regulates GR-GF mapping. For example, the Balinese UV in (1b) and Puyuma CV in (2b) select 2:patient and 2:beneficiary respectively as SUBJ/PIVOT. These UV structures result in a mismatch between GF and GR prominence, informally represented by crossing lines.

- (4) a. Balinese UV in (1b)



- b. Puyuma CV in (2b)

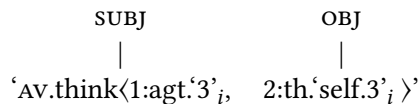


By distinguishing GRS and GFS, we can reflect a prominence mismatch in the non-AV structures in (4) above. This is evident from the interaction between reflexive binding and voice alternation in AN languages. For instance, the AV-UV voice alternation does not affect the acceptability of reflexive binding, exemplified by *awakne* in Balinese (5) and *izipna* in Kavalan (6). For simplicity, the f-structures showing reflexive binding are only given for the examples below.

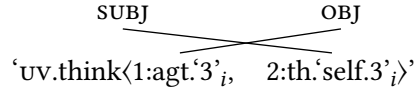
- (5) Balinese (Arka 2003a: 178)

- a. [Ia]_{SUBJ} ngenchang [awakne]_{OBJ}.
 3 AV.think self.3

'(S)he thought of herself/himself.'

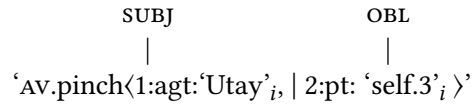


- b. [Awakne]_{SUBJ} kenehang[=a]_{OBJ}.
 self.3 uv.think=3
 ‘(S)he thought of *herself/himself*.’

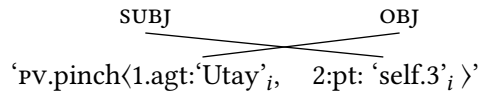


(6) Kavalan (Formosan) (Shen 2005: 57)

- a. Knit [ci Utay]_{SUBJ} [tu izipna]_{OBL}.
 <AV>pinch PN Utay OBL body:3GEN
 ‘Utay pinched at himself.’



- b. Kenit-an=na [ni Utay]_{OBJ} [ya izipna]_{SUBJ}.
 pinch-PV=3ERG ERG Utay ABS body:3GEN
 ‘Utay pinched *himself*.’



c. f-structure of sentence (6a)

PRED	'PINCH<SUBJ OBL>'										
SUBJ	<table> <tr> <td>PRED</td><td>'UTAY'</td></tr> <tr> <td>NTYPE</td><td>PROPER</td></tr> <tr> <td>INDEX</td><td> <table> <tr> <td>PERS</td><td>'3'</td></tr> <tr> <td>NUM</td><td>SG</td></tr> </table> <i>i</i> </td></tr> </table>	PRED	'UTAY'	NTYPE	PROPER	INDEX	<table> <tr> <td>PERS</td><td>'3'</td></tr> <tr> <td>NUM</td><td>SG</td></tr> </table> <i>i</i>	PERS	'3'	NUM	SG
PRED	'UTAY'										
NTYPE	PROPER										
INDEX	<table> <tr> <td>PERS</td><td>'3'</td></tr> <tr> <td>NUM</td><td>SG</td></tr> </table> <i>i</i>	PERS	'3'	NUM	SG						
PERS	'3'										
NUM	SG										
OBL	<table> <tr> <td>PRED</td><td>'PRO'</td></tr> <tr> <td>PRONTYPE</td><td>REFLEXIVE</td></tr> <tr> <td>INDEX</td><td><i>i</i></td></tr> </table>	PRED	'PRO'	PRONTYPE	REFLEXIVE	INDEX	<i>i</i>				
PRED	'PRO'										
PRONTYPE	REFLEXIVE										
INDEX	<i>i</i>										
VOICE-TYPE	AV										

d. f-structure of sentence (6b)

PRED	'PINCH<SUBJ OBJ>'										
SUBJ	<table> <tr> <td>PRED</td><td>'PRO'</td></tr> <tr> <td>PRONTYPE</td><td>REFLEXIVE</td></tr> <tr> <td>INDEX</td><td><i>i</i></td></tr> </table>	PRED	'PRO'	PRONTYPE	REFLEXIVE	INDEX	<i>i</i>				
PRED	'PRO'										
PRONTYPE	REFLEXIVE										
INDEX	<i>i</i>										
OBJ	<table> <tr> <td>PRED</td><td>'UTAY'</td></tr> <tr> <td>NTYPE</td><td>PROPER</td></tr> <tr> <td>INDEX</td><td> <table> <tr> <td>PERS</td><td>'3'</td></tr> <tr> <td>NUM</td><td>SG</td></tr> </table> <i>i</i> </td></tr> </table>	PRED	'UTAY'	NTYPE	PROPER	INDEX	<table> <tr> <td>PERS</td><td>'3'</td></tr> <tr> <td>NUM</td><td>SG</td></tr> </table> <i>i</i>	PERS	'3'	NUM	SG
PRED	'UTAY'										
NTYPE	PROPER										
INDEX	<table> <tr> <td>PERS</td><td>'3'</td></tr> <tr> <td>NUM</td><td>SG</td></tr> </table> <i>i</i>	PERS	'3'	NUM	SG						
PERS	'3'										
NUM	SG										
VOICE-TYPE	PV										

The above data points show that reflexive binding in Balinese and Kavalan takes place at the level of a-structure, as shown in the a-structure representations on the right side. In Balinese, for example, both (5a) and (5b) share the same a-structure, but differ in their respective mapping to GFS: consider the crossing line in (5b) in the UV structure. In Kavalan, both (6a) and (6b) are similar in their a-structure representations except that the non-actor argument in (6a) is non-core, as represented by the vertical line (|). Both languages demonstrate that the reflexive is bound by the subject, as indicated by the subscript *i*. However, the voice alternations trigger a difference in the resulting GFS of the reflexive. It is realised as OBJ in the AV in (5a) and SUBJ in UV in (5b) for Balinese. In Kavalan, on the other hand, the patient is realised as OBL flagged by an OBL marker *tu* in the AV verb in (6a) (due to the ergative system of this language), and it is realised as SUBJ and flagged by the marker *ya* in Patient Voice (PV) in (6b). In both instances, the relationship between the reflexive pronoun (*izipna* ‘body.3GEN’) and its binder (i.e., the intended antecedent) is expressed through coindexation in the f-structure, indicated by the subscript *i*. The PERS and NUM values (i.e., 3SG) of the INDEX attribute of the bindee (*izipna*) are linked or bound to the INDEX values of the binder (*Utay*). This binding of INDEX values between the binder (*Utay*) and the reflexive pronoun (*izipna*) is permissible due to a binding requirement associated with the reflexive pronoun (cf. Rákosi 2023 [this volume]). Crucially, being the first core agent (i.e. <1:agent>) argument, *Utay* outranks the reflexive pronoun (*izipna*) in the a-structure.

The acceptable reflexive binding of SUBJ (i.e. the most privileged argument) by OBJ in (5b) and (6b) would be unexpected if binding took place at the surface grammatical function level because the antecedent (OBJ) has lower syntactic prominence. The occurrence of reflexive binding in non-AV structures confirms the prominence outranking in the a-structure (i.e., $A > P$). This finding emphasizes the necessity of a separate syntacticized a-structure to provide an accurate analysis of reflexive binding phenomena in Austronesian languages, including Balinese and Kavalan.

In LFG, the important characteristics of the AN voice system can be captured using the layered a-structure, and cross-linguistic variation in the voice system is effectively illustrated by the varying transparency of the mapping. A distinction between GFS (the primitives of f-structure) and GRs (the primitives of a-structure) is maintained in western AN, but collapses in accusative languages like English, and Eastern AN languages that lack a symmetrical voice system (see Section 4).

Without this notion of layered structures, the unusual variation in the surface realisation of reflexive binding cannot be easily captured in other theories.⁹

Having shown how the basics of the AN voice system work from an LFG perspective, we now move on to an overview of some typologically interesting phenomena in AN languages in the subsequent sections.

3 Clausal word order

In Western AN languages, there are two broad patterns of clausal word order that are geographically distributed (Blust 2013: 461–461). Verb-initial order is encountered in the AN languages of Taiwan, the Philippines, northern and central Sulawesi, and Madagascar. Philippine-type languages tend to be verb-initial, whilst Indonesian-type languages, including Balinese, Madurese and Indonesian, are verb-medial. Diachronically speaking, the development of these two types of word orders appears motivated primarily by information structure—such as clefting to express contrastive FOCUS (see Section 6)—resulting in synchronic word order variation. The broad classification of Western AN will be used in the ensuing discussion with regard to the typology of word order. However, it should be noted that there is also a great deal of variation across the Philippine-type and Indonesian-type languages due to the flexibility of the order of agent and non-agent arguments relative to the head verb, giving rise to languages with or without a VP and languages with a rigid or flexible subject position (see Riesberg et al. 2019 for further details). In addition, language-internal variation exists, and it has been claimed that some AN languages do not have a fixed basic word order, or that word order choice may differ by voice construction, among other things (cf. Riesberg et al. 2019).

Unlike most Western AN languages, word order varies among Eastern AN languages. The AN languages of eastern Indonesia and many Oceanic languages have typically developed systematic pronominal indexing systems,¹⁰ and therefore show a greater degree of freedom and variation for the ordering of cross-referencing NPs. Thus, there are Eastern AN languages that show SVO clausal

⁹For the sake of brevity, a detailed comparison with other theories is omitted here. We confine our illustration to the use of LFG in analysing AN languages. An in-depth comparative discussion of LFG and other frameworks is provided in Part VII.

¹⁰Some Formosan languages (e.g. Puyuma and Kavalan) have pronominal indexes on verbs. While they closely interact with robust voice verbal morphology, they do not usually contain a complete set of forms exhibiting the full range of case/role alternations. For this reason, we propose that the systematic pronominal indexing systems in Eastern AN languages are distinct from those in Formosan languages.

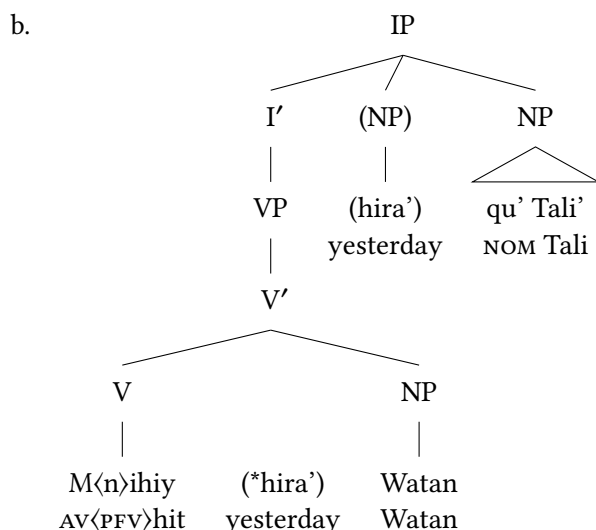
word order with indexed NPs ordered flexibly (e.g. Kambera), and there are others that are verb-initial (e.g. Fijian), and further still, there are other languages which are verb-final as a result of Papuan contact (e.g. Tobati, an AN language spoken in Jayapura Bay, west Papua, and Torau, an Oceanic language spoken in Bougainville; cf. Lynch et al. 2002).

In LFG, word order variation reflects the surface differences between ‘default’ (or unmarked) clausal order and pragmatically marked order. These are dealt with in terms of variation at the level of c-structure (see Andrews 2023 [this volume]).¹¹ Below we illustrate word order variation in Philippine-type, Indonesian-type and indexing type languages, from an LFG perspective.

Verb-initial sentences in Philippine-type languages are finite clause structures with the (inflected verbal) predicate, or the auxiliary, occupying the left-headed inflection (I) node. Hence, a sentence is head (or predicate) initial. However, the precise structures of post-verbal elements vary, with certain languages like Squliq Atayal (Formosan) showing a rigid hierarchical Verb Phrase (VP) structure, whereas others like Tagalog have a non-configurational structure. Evidence for a VP in Atayal comes from an adverbial insertion test. As shown in (7a), *hira* ‘yesterday’ cannot intervene between a transitive verb and its object. The c-structure of (7a) is represented in (7b). Note that LFG adopts a version of X-bar syntax that allows nonbinary branching, as seen in the top/root node of IP in (7b).

- (7) Squliq Atayal (Formosan) (Liu 2017: 41)
- a. M<n>ihiy (*hira’) Watan (hira’) qu’ Tali’
 AV<pfv>hit yesterday Watan yesterday NOM Tali
 ‘Tali hit Watan yesterday.’

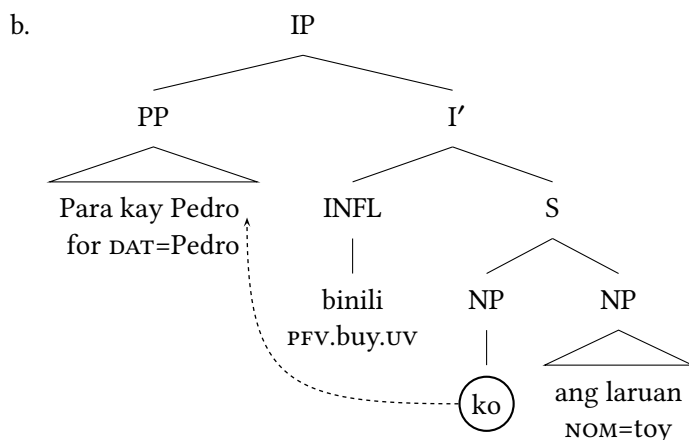
¹¹While c-structure in LFG is modelled using phrase structure trees, with properties possibly following an X-bar schema, it does not represent a deeper ‘universal’ syntactic relation in which, for example, the object or patient argument is uniformly represented in the complement position of a VP as typically characterised by Chomskyan generative models. Further, there is no constituent movement in LFG, even though we may informally refer to ‘fronting’; see Bresnan & Kaplan (1982), Bresnan et al. (2016: chapter 6), Dalrymple et al. (2019: chapter 3), Andrews 2023 [this volume].



Turning to Tagalog, we can posit that the finite sentence (IP) in this language contains a non-configurational (i.e. exocentric, flat) Sentence (S), as shown in (8b) for the example in (8a); for further discussion of exocentricity and the category S, see Andrews 2023 [this volume]. Evidence for this comes from the fact that post-verbal arguments of non-verbal predicates (e.g. SUBJ and OBL) can be freely ordered (Kroeger 1993: 133). There is no surface VP in Tagalog because a second-position (2P) clitic – which must appear in the second syntactic position of a clause in order to obey syntactic-phonological constraints – is hosted by the finite verb alone and not the verb complement if the clause is verb-initial (not exemplified here), or by the first/fronted X(P) as exemplified in (8). Any attempt for a VP (i.e. V and its argument) to host a 2P clitic is ungrammatical (Kroeger 1993: 136).

(8) Tagalog (WMP, Philippines) (Kroeger 1993: 129)

- a. [Para kay=Pedro]=ko binili ang=laruan.
 for DAT=Pedro=1SG.GEN PFV-buy-PV NOM=toy
 ‘For Pedro I bought the toy.’



Variation in predicate-initial word order is pragmatically driven, and allows a unit to be ‘fronted’ to a sentence-initial position before the verb. This position bears a Discourse Function (DF) and is not uniquely associated with a particular grammatical function. In the Tagalog example (8a), the fronted DF, the OBL ‘for Pedro’, structurally occupies the Specifier position of the finite sentence, [Spec,IP], as shown in the c-structure in (8b). While generated under the S node by the phrase structure rule, because =*ko* is a 2P clitic, it is hosted by the Prepositional Phrase (PP), the first syntactic unit, following the final word of the phrase, *Pedro*.¹² Additionally, as in example (9) for Squliq Atayal, the sentence-initial position can be occupied by a grammaticalised topic that co-references SUBJ. This results in a pragmatically marked order for the pseudo-SV(O), namely, SUBJ-VERB-(OBJ/OBL). A pause, indicated by a comma in (9), is observed between the adjoined topic and the IP.

(9) Squliq Atayal (Liu 2017: 202)

Pagay qani (ga), kguh-an na' ngta'.

rice this TOPIC scatter-LV OBL chicken

‘(Speaking of) the rice, (it) was scattered by the chicken.’

¹²Kroeger’s analysis of clitic placement follows the standard approach in LFG (cf. Bresnan et al. 2016: 155), which treats a clitic as a syntactically independent unit like any other word. It occupies a terminal c-structure node, but is post-lexically hosted by another X(P) node due to a prosodic requirement in the syntax-phonology interface in the grammar. A different approach is to treat a clitic as a phrasal affix which does not occupy a terminal syntactic node on its own (cf. O’Connor 2002). See Halpern (1995), Halpern & Zwicky (1996), King (2005), and Bögel et al. (2010), among others, for further discussion of (2P) clitic placement.

The distribution of pronominal special clitics, such as 2P clitics, may also be determined by syntactic-pragmatic conditions that give rise to variations in clausal ordering. This is observed in Pazeh-Kaxabu (Formosan). Pazeh-Kaxabu has two types of bound pronominals: a full set of 2P clitics, and a ‘peripheral’ clause-final clitic. Crucially, the 2P pronominals are strictly used as an operational device so the speaker can direct an addressee’s attention to the predicative element that is syntactically intransitive, as seen in examples (10a–b).

(10) Pazeh-Kaxabu (Formosan) (Li & Tsuchida 2001: 106, 140)

- a. [[Ma-dese_k]_{V:FOCUS-C} [=siw]_{SUBJ}]_{IP}.
STAT-belch =2SG.ABS
‘You *belch!*’ (emphasis added)
- b. [[M<in>e-ken]_{V:FOCUS-C} [=siw]_{SUBJ} sumay=lia]_{IP}?
AV <PFV>eat =2SG.ABS rice(meal)=MODAL
‘Have you *eaten* meals?’ (emphasis added)
- c. [M<in>e-ken asai paj= [isiw]_{SUBJ}]_{IP} ?
AV <PFV>eat what MODAL= 2SG
‘*What* have you eaten?’ (emphasis added)

In (10a–b), the 2P clitic pronoun =siw appears as the sole argument of a simple stative intransitive verb (*madese_k* ‘belch’ in 10a) and an intransitive clause¹³ (*meken* ‘eat’ in actor voice in 10b). These sentences come with an emphatic focus on the predicates¹⁴ (indicated by italicisation in the free translation; cf. 10b and 10c). The free pronoun *isiw* that encodes the SUBJ of a wh-question in sentence (10c) differs from the 2P clitic pronoun in its pragmatic function. Unlike the predicate host in (10b), there is no emphatic focus on the verb in (10c), and meanwhile, the pronominal SUBJs in the two sentences differ in their clausal positions — the free pronoun appears clause-finally, while the 2P clitic appears in an immediately post-verbal position.

Unlike 2P clitics and free pronouns, the host of the peripheral pronominal in Pazeh-Kaxabu is the last word of the clause. The peripheral pronominal clitic, while neutral in case, bears DF for contrastive meaning to encode a highly top-

¹³The issue of semantic versus syntactic transitivity of actor-voice clauses in some AN languages is discussed in Section 4.4.

¹⁴The term “focus” is used in this chapter to refer to the notion in information structure (Zaenen 2023 [this volume]), which is different from the term for the “focus system” that is primarily used by Formosan linguists. The latter will be discussed in Section 4 as “voice alternation.”

ical entity.¹⁵ The use of the peripheral pronominal entails that the post-verbal core arguments are pragmatically ordered according to their DF roles, giving rise to order variation for VOS with a focused SUBJ in (11a) and VSO with a salient referent of OBJ in (11b).

(11) Paze-Kaxabu (Li & Tsuchida 2002: 96)

- a. [Ka-kan-en [nimisiw]_{OBJ} =lia [=aku]_{SUBJ:FOCUS-C}]_{IP}.
 DUR-eat-PV 3ERG =MODAL =1SG.NEUTRAL
 ‘She (the leopard) would surely eat *me*.’ (emphasis added)
- b. [Ta-padudu-i [isiw]_{SUBJ:FOCUS-C} =na [=aku]_{OBJ:TOPIC}]_{IP}.
 HORTATIVE-consult-PV 2SG =MODAL =1SG.NEUTRAL
 ‘Perhaps, let *me* consult you.’ (emphasis added)

Non-predicate-initial Indonesian-type languages, such as Balinese (Arka 2003a), Batak (Erlewine 2018), Madurese (Davies 2010: 249) and Sasak (Wouk 2002), have slightly different structural properties. First, the [Spec,IP] position is occupied by the grammatical SUBJ, accounting for the verb-medial (SVO) structure in these languages. This is exemplified by Balinese in (12) and Madurese in (13).

(12) Balinese (Arka 2003b: 78)

- [[Tiang]_{SUBJ} [[nunas kopi-ne niki]_{VP}]_{I'}]_{IP}.
 1 AV.take coffee-DEF this
 ‘I took this coffee.’

(13) Madurese (WMP, Indonesia) (Davies 2010: 149)

- Sengko’ ng-enom kopi.
 1 AV-drink coffee
 ‘I drink coffee.’

Unlike Tagalog, Indonesian-type languages, such as Toba Batak (Erlewine 2018), Indonesian (Arka & Manning 2008) and Balinese (Arka 2003a), appear to have a VP. Evidence for this comes from constituency tests such as material intervention

¹⁵The 1st person singular pronominal form in Paze-Kaxabu lends empirical support to the emergence and development of split-subjecthood in Formosan languages, where a non-SUBJ agent that bears a high degree of topicality in the discourse is developed to possess syntactic and morphological subject properties. Readers are directed to Liu (2017) for discussion of split-subjecthood.

and joint-fronting. This is particularly evident when the patient/agent argument is indefinite. The material-intervention test is given in (14) for Toba Batak, where a clausal adjunct cannot intervene between a verb and its argument.

- (14) Toba Batak (WMP, Philippines) (Erlewine 2018)
 Man-jaha (*nantoari) buku (nantoari) si Poltak (nantoari).
 AV-read yesterday book yesterday PN Poltak yesterday
 ‘Poltak read a book yesterday.’

The joint-fronting test is evident when the verb receives Contrastive Focus (FOCUS-C) and is required to appear sentence-initially. The whole V+NP string should be included, otherwise the structure is ungrammatical. For example, in contrast to the default SVO order in Balinese in (12), sentence (15a) is a pragmatically marked VOS sentence (as seen from its translation).¹⁶ A postverbal subject is unacceptable, as depicted in (15b).

- (15) Balinese (Arka, own knowledge)
 a. [[Nunas kopi-ne niki]_{VP} , [[tiang]_{SUBJ}]_{IP}.
 AV.take coffee-DEF this (pause) 1
 ‘Taking this coffee was what I did.’
 b. *[[Nunas]_{V:FOCUS-C}, [[tiang]_{SUBJ} [kopi-ne niki]_{VP}]_{IP}.

Clause structure variation in Indonesian-type languages is usually driven by pragmatic considerations, primarily to express varying levels of informational salience or attention, for example, emphatic or contrastive focus and frame setting or topic (Arka 2003a: 257–260, Arka & Sedeng 2018, Davies 2010: 175–176, Norwood 2002: 104–107). The unit that functionally bears a Contrastive Discourse Function is fronted sentence-initially. Following Arka (2021), we explicitly represent contrastive FOCUS and TOPIC as FOCUS-C and TOPIC-C, respectively, where necessary.

In order to integrate the latest advancements in the study of information structure within Austronesian languages (Riesberg et al. 2018) and beyond (Dalrymple & Nikolaeva 2011, Zaenen 2023 [this volume], among others), we deviate slightly from the LFG representation of TOPIC and FOCUS proposed by, for example, Bresnan & Mchombo (1987). Our approach introduces distinct types of Discourse

¹⁶VOS order is also possible when the subject is an afterthought TOPIC. This is a different structure, and the pragmatics and related prosody are different.

Functions, including FOCUS-C, beyond the traditional analysis assumed in LFG during the 1980s and early 1990s. Apart from contrastive TOPIC and FOCUS, the fine-grained realm of TOPIC and FOCUS encompasses additional types such as 'new, first mentioned TOPIC', 'default TOPIC', 'secondary TOPIC', and 'new/completive FOCUS'. Arka & Sedeng (2018) provide examples of these categories in Sembiran Balinese. The suggested contrastive DF is ideally situated within an independent i-structure (King 1997, Andréasson 2007, Butt 2014, among others), although it can also be, for simplicity, integrated within LFG's conventional unified f-structure representation (cf. (16c) below). The FOCUS-C case is exemplified by (15) in Balinese above and by (16) in Indonesian below.

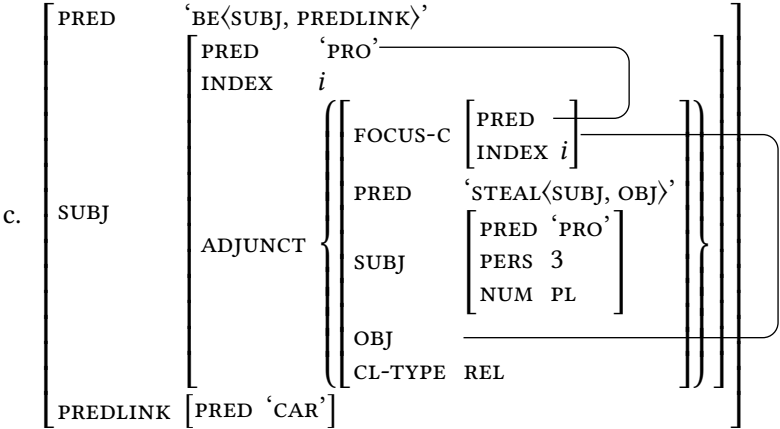
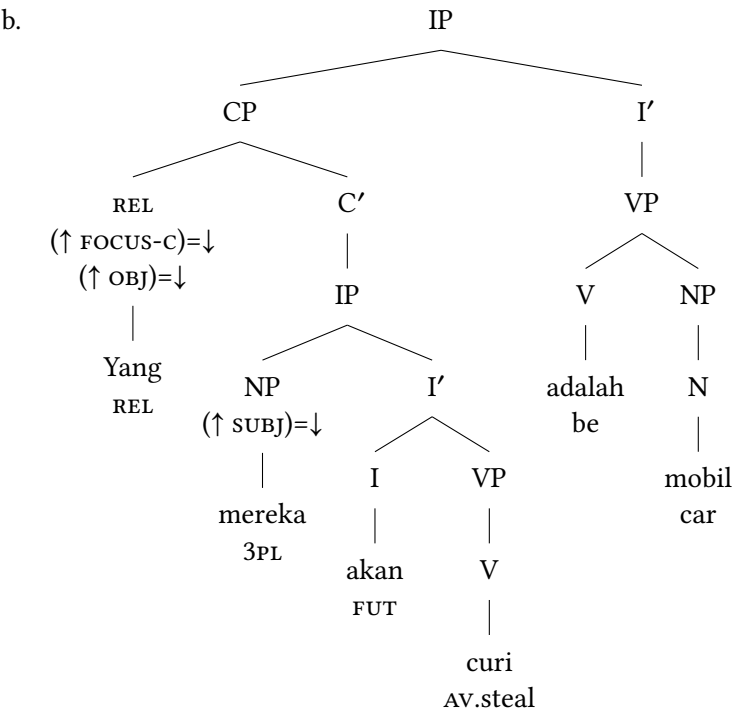
However, the precise structural position of contrastive DFs may vary depending on whether or not a language has a functional complementiser (C) word. For a language like Indonesian, which has a C (*bahwa* 'that'), the contrastive DF is in [Spec,CP]. That is, a finite clause is CP, the maximal projection of C. The finite CP in Indonesian is evident in the relative clause (RC) with *yang* bearing FOCUS-C, as exemplified in (16a) (cf. Arka 2011: 78-80 and Arka (2021) for details). The c-structure tree is given in (16b), showing that the pronominal relativiser *yang* is grammatically OBJ, and the RC is structurally OSV. The f-structure is shown in (16c).¹⁷

(16) Colloquial Indonesian¹⁸ (WMP, Indonesia) (Arka, own knowledge)

- a. [Yang mereka akan curi]_{CP} adalah mobil.
 REL 3PL FUT AV.steal be car
 'The thing that they were going to steal was a car.'

¹⁷Note that the Indonesian copular verb *adalah* is analyzed as requiring PREDLINK, which is one way of analyzing a nominal predicate in LFG. For discussion of single-tier/double-tier analysis of non-verbal predicates, see Andrews (1982), Butt et al. (1999), Dalrymple et al. (2004), among others.

¹⁸Standard Indonesian and Colloquial Indonesian differ in their morphological properties of verbs and the formation of relativisation. See Arka (2021) for more exemplification.



Unlike Indonesian, Balinese has no complementiser C equivalent to English *that* (Arka 2003a).¹⁹ A fronted element bearing a contrastive DF can be analysed as being left-adjoined to IP. This structure is shown by the IP subscripts in example

¹⁹However, certain prepositions (e.g. *unduk* ‘about’) and conjunctions (e.g. *apang* ‘so that’) may function like complementisers in particular contexts (Natarina 2018: 54).

(15a). Note that the fronted element bearing a marked FOCUS-c is typically given stress with a clear pause after it (indicated by a comma above) resulting in a VOS structure.

Like Tagalog, the clausal linear order in Indonesian-type languages may also vary if the SUBJ is a 2P clitic. The variation may involve contrastive DFs. For instance, Sasak has a 2P clitic SUBJ (Austin 2004) which appears after the first constituent for independent syntactic-phonological reasons, giving rise to clausal word-order variation. Thus, while S-(Auxiliary)-V-O is the unmarked order in Sasak, the subject may also be cliticised to an auxiliary if it is the first word in the sentence; this results in an Aux-S-V order, as seen in (17a). In (17b), however, the verb is fronted sentence-initially, as it bears FOCUS-c. Therefore, it hosts the subject clitic and results in a VSO order.

(17) Ngenó-ngené Sasak (WMP, Indonesia) (Asikin-Garmager 2017: 29, 32)

- a. Kenyengken=ne tokol.
 PROG=3 sit
 ‘They (the women) were sitting.’
- b. M-pantòk=ne_i begang inó (isiq lóq Mus_i).
 PREDFOC-hit=3 rat that by ART.M Mus
 ‘Mus hit the rat. (He finally got it!)’ (emphasis added)

In contrast, the clausal word order is typically fixed when an argument is generic or indefinite (see Section 4.4 for further discussion of definiteness). For example, the Balinese generic statement about a cow in (18a) must be in SVO; a VOS variant is unacceptable, as in (18b).

(18) Balinese (Arka 2019: 261)

- a. Sampi ngamah padang.
 cow AV.eat grass
- b. *Ngamah padang sampi.
 AV.eat grass cow
 ‘A cow eats grass.’

Some AN languages in the peripheral regions, geographically distant from their original homeland of Taiwan, are morphologically isolating and typically exhibit rigid SVO clause order. These languages are encountered on Flores Island in Indonesia and other peripheral areas, such as in Southeast Asia and the Pacific. Structurally, their clauses are like Indonesian-type languages with good evidence for a surface VP. Consider the following intervention test in Rongga (central Flores), a highly isolating language, where the verb and object form a VP:

(19) Rongga (CEMP, Indonesia) (Arka 2016: 192)

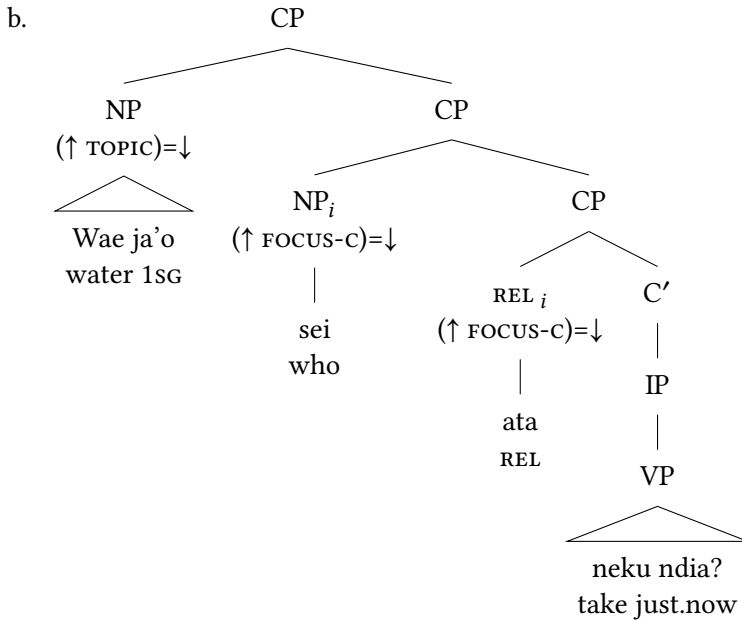
- a. Ardi [ngedho wolo]_{VP} **nembumai**.
 Ardi see mountain yesterday
 ‘Ardi saw a/the mountain.’
- b. *Ardi ngedho **nembumai** wolo.

Rongga has developed a relativiser (REL) from the noun meaning ‘person’, *ata*, exemplified in (20a). As in Indonesian, it may also be analysed structurally as appearing in [Spec,CP], bearing a contrastive DF (Arka 2016). The c-structure is given in (20b), which shows that sentence (20a) is a highly marked structure. Both TOPIC and FOC are present, with TOPIC preceding FOC in the left periphery. Note that FOCUS-C in this example is associated with two elements having the same referent (indicated by the subscript *i*). Hence, it is doubly marked: first by the relativiser, *ata*, and second by the fronted question word (Q), *sei* ‘who’. The sentence is a cleft structure with the Q, *sei*, being the (fronted) predicate and the relative clause being the SUBJ (as shown by the literal translation). Considering that relativization introduces a contrasting emphasis by focusing on or restricting a specific referent under discussion or question, we analyze the relativizer as carrying FOCUS-C. In example (20a), for instance, multiple individuals were present, and the relative clause singles out one of them through the event of ‘taking (my) water’.²⁰

(20) Rongga (Arka 2016: 212)

- a. Wae ja’o, sei ata neku ndia?
 water 1SG who REL take just.now
 ‘As for my water, who’s the one taking (it) just now?’
 (Lit. ‘As for my water, the one taking (it) just now is who?’)

²⁰It should be noted that, from the broader viewpoint of the matrix noun phrase, the relativizer is linked to the specific referent being talked about, to which the relative clause adds its semantic restriction. Therefore the relativizer can also be analyzed as a topic (cf. Bresnan & Mchombo 1987).



Unlike Philippine-type and Indonesian-type languages, the indexing-type AN languages of eastern Indonesia and Oceania have developed systematic pronominal indexing systems. The salient grammatical trait of AN (symmetrical) voice is either disappearing or already lost in these languages. As a result, these languages have relatively free word order determined largely by discourse pragmatics.

For example, Kambera in Sumba, eastern Indonesia, has developed different sets of bound pronominal indexes (NOM, ACC, GEN and DAT) that appear on the predicate with a fixed order (Klamer 1998: 79). In example (21) below, *na-* and *-nya* are subject and object arguments, respectively. They appear with free cross-referencing NPs. These free NPs are optional and freely ordered, hence allowing NP_i -[SUBJ_{*i*}-V-OBJ_{*j*}]=NP_{*j*} (or SVO: (a)) and NP_j -[SUBJ_{*i*}-V-OBJ_{*j*}]=NP_{*i*} (or OVS: (b)) orders. OSV, despite not being shown here, is also possible. The SVO structure in (21a) is the default/unmarked order for transitive clauses, and OVS is a marked order when OBJ is contrastive TOPIC (Klamer 1996: 22). The basic word order for an intransitive sentence is, however, VS (Klamer 1998: 85). Kambera syntax is typologically like Chicheŵa (albeit with a difference in the ‘agreement’ status

of the verbal SUBJ marker),²¹ and it can be analysed in LFG in the same way as outlined in Bresnan & Mchombo (1987): the bound pronominal indexes are the actual syntactic arguments whereas the free NPs bear DFs, and are pragmatically linked to the arguments, which gives rise to some kind of anaphoric agreement.

(21) Kambera (CEMP, Indonesia) (Klamer 1996: 13)

- a. Ka nyuna_j na_j-tinu-nya_k na lau_k.
 CNJ she 3SG.NOM-weave-3SG.DAT ART sarong
 ‘So that she weaves the sarong.’ (Lit. ‘she she-weaves-it the sarong.’)
- b. Ka na lau_k na_j-tinu-nya_k nyuna_j.
 CNJ ART sarong 3SG.NOM-weave-3SG.DAT she
 ‘So that the sarong was woven (by her).’ (Lit. ‘the sarong
 she-weaves-it she.’)

To sum up this section, LFG is well-suited for analysing word order variation across different types of AN languages based on a few parameters that are empirically motivated (e.g. VP vs non-configurational, head-initial vs head-final, a contrastive DF in [Spec,CP], [Spec,IP] or left-adjoined to IP). This is made possible by the c-structure representation in LFG which follows a flexible version of X-bar Theory, and which not only captures cross-linguistic structural similarities (e.g. headedness in lexical and functional categories), but also varying language-specific properties (e.g. the distinction between the endocentric phrase and an exocentric S that is not X'-theoretic, and their multiple branching units).

²¹The verbal SUBJ/OBJ markers in Chicheŵa differ from those in Kambera in the following ways. As in Kambera, the SUBJ marker in Chicheŵa is obligatory in the verbal template. However, the Chicheŵa SUBJ marker is only optionally pronominal. It serves as the actual argument when there is no free SUBJ NP. Therefore, unlike in Kambera, it can also function as a ‘syntactic’ agreement marker when there is a free SUBJ NP present. The SUBJ/OBJ markers in Kambera hold a compulsory position within the verbal template and are consistently pronominals, meaning they refer to entities even in the absence of their corresponding free NPs. Consequently, these markers do not serve as syntactic agreement markers. In this regard, these affixes share similarities with verbal affixes found in Papuan languages (Arka et al. forthcoming) and certain Australian Aboriginal languages like Wambaya and Warlpiri. In these languages, the affixes exhibit ambiguity as they can function both as (anaphoric) agreement markers and incorporated pronominals (see Austin & Bresnan 1996 and references therein).

4 Grammatical functions and alternative argument realisations

4.1 Introduction

In LFG, grammatical functions are dealt with independently from the a-structure. Recall that in Section 2, we briefly introduced the basics of the voice system in Western AN and the rationale behind adopting a syntacticised a-structure in LFG (following Manning 1994, Arka 2003a, Arka & Manning 2008). Certain aspects of our architecture and related representations/mechanisms differ slightly from the assumptions generally adopted in LFG. One example is the argument linking/mapping mechanism (cf. Findlay et al. 2023 [this volume]). The presentation used in this chapter is to account for salient symmetrical AN voice system where both accusative and ergative properties are observed within the same grammatical system, as in Balinese (Arka 2003a), which allows an underlying argument to have alternative GF realisations, like the a-subject/object to be the surface OBJ/SUBJ as seen in Balinese in (5) and Kavalan in (6). These languages are classified as alternating languages for our discussion here.

On the other hand, AN languages of the indexing type, like Kambera, Woi, and Taba, lack symmetrical voice and the associated SUBJ/PIVOT distinction, and thus tend to be non-alternating languages (Klamer 1996, Bowden 2001, Sawaki 2016). Typically, their AN voice morphology and related voice system have disappeared. Consequently, core arguments (subject/object) do not have surface GF alternations like the kind witnessed in the alternating languages. The transitive subject and object are consistently surface SUBJ and OBJ, respectively.

In other words, non-alternating languages tend to have fixed argument linking. In a genuinely non-alternating system, there is typically no distinction between GFS and GRS. This gives rise to the salient typological property of non-alternating systems that GFS are typically semantically transparent. For a transitive predicate, SUBJ is therefore always the most agent-like argument as seen, for example, in Kambera in the examples in (21) above. The bound (NOM) proclitic *na=* is always the ARG1:agent/SUBJ argument of a transitive verb in this language, even when it is cross-referenced by a postposed free NP as in (21b). That is, sentence (21b) is not grammatically passive despite being given a passive translation in English; the agent is neither OBL nor an adjunct (cf. the pronominal marking in a verbal cluster in Klamer 1996). Given the semantic transparency of GFS, intransitive predicates unsurprisingly show a split-S property in non-alternating AN languages. This is seen in, for example, Acehnese in examples (45)-(46) below.

In the ensuing sub-sections, we present, from the LFG perspective, how GFs (e.g. SUBJ, OBJ, OBL) are realised differently in the languages that have robust voice systems (i.e. alternating languages) and in those that do not (i.e. non-alternating languages). Specific diagnostics to identify SUBJ, OBJ and OBL, or their grouping as core versus non-core arguments, vary depending on the language type and morphosyntactic resources available (such as verbal morphology, pronominal marking, and phrase/case marking) in a given language. The complexity of the properties has led to a wide variety of competing analyses, for example, in the context of grammatical alignment systems to be discussed in Section 4.4. We begin by clarifying the subtle and crucial difference between SUBJ and PIVOT.

4.2 SUBJECT and PIVOT

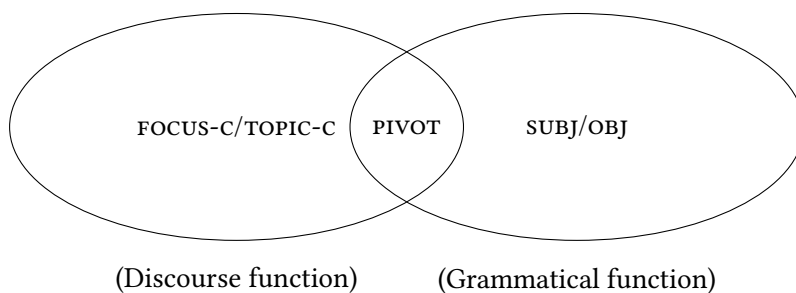
There have been competing proposals within LFG for analysing and representing the predicate's most prominent argument, traditionally referred to as 'subject' (cf. subjecthood in Falk 2006, Belyaev 2023a [this volume]). In this book chapter, we keep the standard LFG conception of SUBJ(ECT) (i.e. in upper case): it is the surface grammatical subject, the most prominent GF on the relational GF hierarchy. It is part of *f*-structure, distinct from the thematic subject (or $\hat{\theta}$), the most prominent role on the thematic hierarchy, and part of the semantic argument structure (Bresnan & Kanerva 1989, 1992).²² It is also distinct from the *a*-subject, the syntacticised *a*-structure subject (Manning 1994). Separating SUBJ from *a*-subject is necessary to account for symmetrical voice alternations and related properties in AN languages (cf. the separation of GF-GR in Section 2), otherwise certain unusual phenomena, such as the binding of SUBJ by OBJ (e.g. in Balinese example [5]) cannot be accounted for.

Furthermore, it is essential to distinguish SUBJ from PIVOT to account for the complex morphosyntax/pragmatics interface, which constrains voice alternation in certain constructions such as fronted questions in Balinese and Amis (to be discussed in Section 4.2.2). We also want to emphasize our conception of PIVOT as schematised in (22a): PIVOT is at the interface of syntax and pragmatics; that is, it shows grammatical properties (i.e. GF-related, typically intersecting with SUBJ, though not always) as well as discourse-pragmatic (DF) properties (e.g. FOCUS-c). In this sense, PIVOT is an 'overlay' or 'intersection' of GFs and DFs (Arka 2021). PIVOT is evident in the formation of bi-clausal structures, such as relativisation and coordination, and other grammatical mono-clausal structures involving

²²The thematic subject roughly corresponds to the so-called 'logical subject', or the most prominent role-based A/S in linguistic typology (i.e. agentive argument of transitive verb, or sole argument of intransitive verb; see Jespersen (1924) and Manning (1994: 7).

marked DFS, such as FOCUS-C in Balinese fronted questions. Informally, licensing SUBJ to bear FOCUS-C as PIVOT can be represented as (22b), thus [DF-C=SUBJ] PIVOT for SUBJ PIVOT.

(22) a.



b. [FOCUS-C=SUBJ]_{PIVOT}

The notion of ‘pivot’ has been discussed and used in previous typological research. In what follows, we briefly provide some context for the conception of PIVOT adopted in this chapter. Its usage here is broadly aligned to descriptive functional and typological linguistics in place of ‘Subject’ and ‘Topic’ in the analysis of ergativity (Chao 1968, Heath 1975, Dixon 1979, Foley & Van Valin 1984, Van Valin & LaPolla 1997). The explicit incorporation of PIVOT into LFG was proposed by Manning (1994) to replace the GF attribute label SUBJ in f-structure (or Manning’s GR-structure) and also to account for ‘inverse’ mapping in ergative languages while maintaining LFG’s ability to account for ‘straight-through’ mapping in familiar accusative languages. In short, Manning’s PIVOT is intersubstitutable with the standard LFG’s SUBJ to capture the cross-linguistic variation and similarity of ergative and accusative systems. Like SUBJ, PIVOT in Manning’s proposal is a subcategorised GF that is licensed by the head PRED. This is an important point that makes Manning’s PIVOT different from Falk’s proposal, to which we now turn.

Falk (2000, 2006) also incorporates PIVOT into LFG. While his conception of PIVOT is broadly in line with PIVOT in typological/functional linguistics (Dixon 1979, Foley & Van Valin 1984, Dixon 1994) and with PIVOT in Manning’s proposal, Falk’s PIVOT in LFG is different in the following respects. First, Falk’s PIVOT is a slightly narrower notion than the generally understood PIVOT in language typology, and in Manning’s interpretation. It is only related to what Schachter (1977) calls reference-related properties of subject, not role-related ones. That is,

in Falk's conception, *PIVOT* is a syntactic function primarily for cross-clausal, combinatoric purposes (Falk 2006: 76).²³

Second, given that it is a syntactic function, like in Manning's conception, Falk's *PIVOT* is an attribute in the *f*-structure. However, it should be noted that there is an element of grammaticalisation of topichood in clause combining processes. For example, the zero or unexpressed argument in control structures is strongly motivated by topicality and pragmatic efficiencies in cognitive processing (Givón 2001: 219; Hawkins 2004: 163-165). Thus, our conception of *PIVOT* as schematised in (22) is slightly different from Falk's in that it is not purely syntactic. *PIVOT* should also be understood as carrying a (grammaticalised) element of discourse-pragmatics in the interface with syntax.

Third, the crucial difference between Falk's and Manning's proposals relates to the status of *PIVOT* in relation to the deeper conception of argument structure. Falk's *PIVOT* is more like *DFS* or *ADJUNCTS* in that it is not part of the *PRED*'s argument structure. In contrast, Manning's *PIVOT* is like *SUBJ* in that it is licensed by the predicate argument structure. There is good evidence that *PIVOT* is grammatically constrained due to its tight link to the *PRED*'s argument structure. For instance, *PIVOT* selection in relativisation and fronted *QS* in Balinese impose a verbal voice constraint. Such a constraint is unexpected on Falk's conception of *PIVOT* as a non-subcategorised or adjunct-like *GF*. For this reason, our conception of *PIVOT* is in line with Manning's interpretation rather than Falk's. Our *PIVOT* is also in agreement with the widely used notion of *pivot* in typological linguistics.

Finally, it is worthwhile briefly commenting on Falk's conception of *PIVOT* and \hat{GF} , and their related mapping. The notation of \hat{GF} (parallel to $\hat{\theta}$ in thematic structure) means the highest *GF* in the subcategorisation frame of the head *PRED*. Since there is no syntacticised *a*-structure (distinct from *f*-structure) in Falk's framework, his \hat{GF} is equivalent to the conflated *SUBJ* in the traditional LFG analysis of *GFS* (cf. Bresnan & Kanerva 1989), and Manning's syntacticised *a*-structure. Crucially, the *GF-PIVOT* mapping in Falk's analysis does not result in *GF* alternations. For example, unlike in our analysis where the *AV-UV* alternation changes the mapping of agent and patient, which results in the patient being mapped onto *SUBJ* in *UV*, the *UV* structure in Falk's analysis keeps the patient as syntactic *OBJ* and the agent as \hat{GF} (i.e. his *SUBJ*). This is surprising and not empirically supported: the patient of the *UV* in Balinese shows up in the surface syntax as

²³Falk's (2006) conception of *PIVOT* as a syntactic function has been extended in Falk (2007) to account for pragmatic-semantic information in NP syntax (i.e. construct state nominals (CSN) in Hebrew: cf. Sadler 2023 [this volume]). The function of *PIVOT* in AN languages differs from the CSN in Hebrew in its application at the clausal level, where it operates exclusively in the symmetrical voice systems, and is most evident in clause combining.

grammatical SUBJ, not OBJ. The evidence comes from the fact that the patient is structurally in the preverbal SUBJ position. In contrast, the agent (which would be SUBJ in AV) appears as OBJ in UV, appearing in the post verbal position (cf. examples 1a–b).

To conclude, our conception of PIVOT is more in line with Manning’s interpretation than Falk’s interpretation. However, unlike Manning’s proposal, we keep the standard LFG conception of SUBJ in f-structure, as we want to keep SUBJ as the clause-internal and most prominent GF, licensed by the head PRED. This is the SUBJ in its role-related dimension in connection to the PRED, distinct from PIVOT (which encapsulates its other clause-external reference-related dimension; Schachter 1977). In addition, and unlike in Falk’s and Manning’s proposals, we do not represent PIVOT as a separate attribute in f-structure, given the nature of PIVOT with overlapping GF-DF properties as shown in (22a). Its presence can be captured as a construction-type (or language-specific) constraint: see Section 6.3.

In what follows, we discuss and exemplify AN SUBJ and PIVOT further. We begin by illustrating the major differences among AN in the morphosyntactic and behavioural properties of clause-internal SUBJ, and then move on to cases where PIVOT is also present. While SUBJ and PIVOT are oftentimes the same argument, they may diverge (Arka 2021).

4.2.1 SUBJECT: Voice marking and argument flagging

Voice marking encodes SUBJ selection. There are at least three types of voice marking across the AN languages: (i) a multi-way voice system without distinct passive/applicative morphology; (ii) a two-way (AV versus UV) voice system, typically with distinct passive/applicative morphology, and; (iii) a restricted and mixed voice-indexing system. Each is discussed below, including its related argument flagging, from an LFG perspective.

SUBJ selection in multi-way voice systems is encountered in Formosan/Philippine-type languages such as Puyuma (Teng 2008), Tagalog, Kelabit (Borneo), Talaud (North Sulawesi; see Utsumi 2013) and Malagasy. The systems in these languages exhibit several salient properties. First, verbal morphology selects SUBJ as having a specific semantic role rather than a generalised role. This role-specific linking of SUBJ is particularly clear in non-actor voice types. Tagalog, for example, shows Patient Voice (PV), Locative Voice (LV), Instrumental Voice (IV), and Dative Voice (DV, including dative/goal/benefactive), in addition to Actor Voice (AV: Foley & Van Valin 1984: 135; Arka 2003b). Kelabit, on the other hand, shows a simpler system with a three-way opposition between AV, UV, and IV. For simplicity, only AV-PV-LV alternations, like in Tagalog, and AV-PV alternations, like in Kelabit, are given in (23)–(24) below.

(23) Tagalog (Kroeger 1993: 14)

- a. M⟨um⟩ili **ang**=lalake ng=isda sa=tindahan.
 ⟨PFV.AV⟩buy SUBJ=man CORE=fish NONCORE=store
 ‘The man bought fish at the store.’

SUBJ OBJ_{PATIENT}
 ‘AV.buy⟨1:agt, 2:pt ⟩’

- b. B⟨in⟩ili-Ø ng=lalake **ang**=isda sa=tindahan.
 ⟨PFV⟩buy-PV CORE=man SUBJ=fish NONCORE=store
 ‘The man bought the fish at the store.’

OBJ_{AGENT} SUBJ
 ‘pv.buy⟨ 1:agt, 2:pt ⟩’

- c. B⟨in⟩ilih-an ng=lalake ng=isda **ang**=tindahan.
 ⟨PFV⟩buy-LV CORE=man CORE=fish SUBJ=store
 ‘The man bought fish at the store.’

OBJ_{AGENT} OBJ_{PATIENT} SUBJ
 ‘LV.buy⟨ 1:agt, 2:pt, 3:loc ⟩’

(24) Kelabit (WMP, Indonesia) (Hemmings 2021: 161)

- a. **La’ih sineh** nenekul nuba’ ngen seduk.
 man DEM AV.PFV.spoon.up rice with spoon
 ‘The man spooned up his rice with a spoon.’

SUBJ OBJ
 ‘AV.spooned.up⟨1:agt,2:pt⟩’

- b. **Nuba’ sikul** la’ih sineh ngen seduk.
 rice PV.PFV.spoon.up man DEM with spoon
 ‘That man spooned up rice with a spoon.’

OBJ SUBJ
 ‘pv.spooned.up⟨1:agt,2:pt⟩’

SUBJ selection is also indicated by structural properties, such as syntactic position and flagging. In Tagalog, SUBJ is flagged by *ang*= in (23) above.²⁴ In Kelabit,

²⁴The intransitive ⟨1:agt⟩ (or S) argument in Tagalog is also flagged by *ang*=, providing robust evidence for clause-internal subjecthood (i.e. the sole core intransitive argument is SUBJ):

(i) Nagsalita *ang*=babae.
 spoke SUBJ=woman
 ‘The woman spoke.’ (De Guzman 1988: 323-324)

SUBJ is a bare NP that occurs preverbally, and has no prepositional flagging to distinguish it from OBL.

Second, the data points above exemplify the hallmarks of the AN symmetrical voice system in two respects: morphologically and syntactically (Foley 1998, Arka 2003b, Himmelmann 2005, Riesberg 2014). In terms of morphological marking, all voice types are equally marked, as clearly seen in Tagalog and Formosan languages, such as Puyuma (Teng 2008) and Pazeh-Kaxabu (Yeh 2019). None of their voice marking is morphologically ‘default’. As for Kelabit, the root of the PV verb is *sikul*, and the PV marking involves *i*-ablaut and sibilization of /t/ to /s/, analysable as a variant of the infix *in-* also seen in Tagalog.²⁵ In LFG, semantically transitive predicates, such as ‘buy’ in (23) and ‘spoon up’ in (24), are listed in their lexical entries as verbal roots with a-structures containing ⟨1:agt, 2:pt⟩ (i.e. the most actor-like and patient-like arguments are the first two ordered core arguments). Voice morphology is a marker for SUBJ linking, following general principles in Lexical Mapping Theory (LMT), and will be further discussed in Section 4.3.

Another syntactic hallmark of the AN symmetrical voice system is that core arguments are equally selectable as SUBJ without obligatory demotion of any other core argument in the argument structure. This results in a non-AV alternation with cross-linking as depicted in (4), where ⟨1:agt⟩ remains the most prominent argument. Evidence for the non-demotion of ⟨1:agt⟩ comes from reflexive binding as demonstrated in Balinese (5) and Kavalan (6). Other evidence comes from argument marking/flagging. This is clearly demonstrated in the AV-PV alternation in Tagalog in (23a–b). The alternative linking between ⟨1:agt⟩ and ⟨2:pt⟩ to SUBJ and OBJ correlates with the alternative flagging with *ang=* and *ng=*. The phrase markers *ang=* and *ng=* in Tagalog flag SUBJ and OBJ respectively. Hence, in the PV in (23b), ⟨1:agt⟩ remains core as it is flagged with *ng=*. This non-demotion property is what typologically distinguishes the AN symmetrical voice system from Indo-European languages like English.

Next, SUBJ selection in AN languages with two-way voice systems is typically encountered in the Indonesian-type. It shows similar symmetrical voice properties to those observed in Tagalog with the exception that the selection of a peripheral semantic role as SUBJ requires a specific applicative marker. Consider the Balinese examples in (25b)–(25c) below, which is a near equivalent of the LV in Tagalog, as seen previously in (23c):

²⁵We thank an anonymous reviewer for pointing this out.

(25) Balinese (Arka 2014b: 60, 75)

a. Ia meli baas (sig dagang-e ento).

3 AV.buy rice at trader-DEF that

‘(S)he bought rice from the trader.’

SUBJ OBJ OBL

‘AV.buy<1:agt, 2:th | 3:loc/source>’

b. Ia meli-nin dagang-e ento baas.

3 AV.buy-APPL trader-DEF that rice

‘(S)he bought rice from the trader.’

SUBJ OBJ OBJ_{THEME}

‘AV.buy<1:agt, 2:loc/source, 3:th >’

c. Anak-e nto belin-in tiang potlot.

person-DEF that UV.buy-APPL 1 pencil

‘I bought a pencil from the person.’

OBJ SUBJ OBJ_{THEME}

‘uv.buy<1:agt, 2:loc/source, 3:th >’

In Balinese, the two-place transitive verb *beli* ‘buy’ obligatorily requires the applicative marker *-(n)in* in order to add a locative/source to the base structure as a core argument. Compare the locative/source role flagged by *sig*²⁶ in (25a) and the unflagged locative/source argument in (25b). The latter is licensed by the verb that contains the applicative morpheme *-in*, and receives a P(atient)-like core status, resulting in a ditransitive construction. Crucially, with an applicative verb (25b–c), the locative/source argument is promoted to the second most prominent position among the core arguments (i.e. <2:loc/source>), essential for its selection as SUBJ; hence, it can appear sentence-initially without flagging as shown in (25c). Similar to Formosan/Philippine languages, core arguments are equally selectable as SUBJ in two-way voice systems (i.e. evidencing the symmetry of syntax), except that the latter languages require a distinct applicative marker, while the former have more robust verbal voice morphology.

Additionally, AN languages of the Indonesian type often have a real passive voice. Sundanese, for example, has a passive marked by *di-*. In passive voice, <1:agt> is demoted to non-core status, resulting in the promotion of patient to the first argument and its link to SUBJ, as shown in (26b).

²⁶The noun phrase flagged by *sig* in (25a) differs from other non-thematic locatives of OBL adjunct in terms of its thematic animacy (versus inanimate location marked by *ka*). See Arka (2014b) for other syntactic properties targeting the distinction between arguments and adjuncts.

(26) Sundanese (WMP, Indonesia) (Davies & Kurniawan 2013: 123)

a. **Asép ngirim** buku ka Enéng.

Asep AV.send book to Eneng

‘Asep sent a book to Eneng.’

SUBJ OBJ OBL

‘AV.send<1:agt, 2:pt | 3:go>’

b. **Buku éta di-kirim** ka Enéng ku Asép.

book that PASS-send to Eneng by Asep

‘The book was sent to Eneng by Asep.’

SUBJ OBL OBL

‘PASSV.send<1:pt | 2:agt, 3:go>’

AN languages of the indexing type, such as Kambera, Kodi and Wooi, also show clause-internal evidence for SUBJ even in the absence of a typical voice system. In these languages, SUBJ is expressed by a pronominal index (clitic/affix) that commonly exhibits a NOM pattern. In Wooi, for example, the verbal prefix *he-* indexes the intransitive SUBJ in (27a) and transitive SUBJ in (27b). Incidentally, a free NP would optionally cross-reference the SUBJ index for pragmatic reasons (e.g. to express contrastive FOCUS) or for semantic reasons (e.g. to express an associative plural as seen in 27b). In our LFG analysis, the index *he-* fills the SUBJ slot in the verbal template. Since *he-* is referential, it contributes [PRED=‘PRO’], [NUM=PL], and [PERS=3] to the value of SUBJ.

(27) Wooi (CEMP, Eastern-Indonesia) (Sawaki 2016: 203, 206)

a. Henda.

he-t-ra

3PL-PL-go

‘They went.’

b. Jon hendora Agus hia na ramdempe.

Jon **he-t-rora** **Agus hia** na ramdempe

John 3PL-PL-hit Agus 3PL LOC yesterday

‘John and associates hit Agus and associates yesterday.’

4.2.2 Behavioural properties of SUBJECT/PIVOT

In the introduction to Section 4, we clarified the theoretical orientation for the terminology used here to denote the distinct notions of SUBJ and PIVOT. In this section, we focus specifically on how PIVOT is motivated, and exemplify cases in AN languages where PIVOT must be strictly identified with SUBJ (henceforth, SUBJ/PIVOT) and other cases where PIVOT is not necessarily SUBJ. We begin with SUBJ/PIVOT cases.

Evidence for the SUBJ/PIVOT constraint is observed in the fronted Qs in Balinese and Amis. Consider the Qs *apa* ‘what’ in Balinese (28a) and *cima* in Amis (29) below. They appear *in situ* because they are simply ‘weak’ FOCUS. By contrast, when the Qs are placed sentence-initially (i.e. fronted) as in (28b) and (30), they must be understood as SUBJ. This discourse prominent property of FOCUS-C has turned SUBJ into the highly privileged status of SUBJ/PIVOT that is borne by the fronted Q. Now consider the contrasting status of arguments that do not involve the overlay function of PIVOT. In the Balinese example (28a), the Q is not fronted. Even for the fronted Q in example (28b), the sentence is only acceptable with reading (i) (indicated by the solid line). While the NP *ci* is closer to the subject position, it can only be understood as OBJ.

(28) Balinese (Arka 2003a: 27)

- a. [Ci ngalih **apa** ditu ibi]_{IP}?
 2 AV.search OBJ there yesterday
 ‘What did you look for there yesterday?’ (in-situ Q=OBJ_{TH})
- b. **Apa** ci [___ ngalih ___ ditu ibi]_{IP}?
 [what]_{FOCUS-C} 2 SUBJ AV.search OBJ there yesterday
 (PIVOT)
 i) ‘What looked for you there yesterday?’ (fronted Q=SUBJ/PIVOT.agt)
 (e.g. a ghost might have disturbed the addressee)
 ii) NOT FOR ‘What did you look for there yesterday?’

(29) In-situ Q (Central Amis - Formosan) (Yeh, fieldwork data)

- a. [Mi-palo’-ay [**cima**]_{SUBJ/FOC} ci Mayaw-an]_{IP}?
 AV-whip-REAL who PN Mayaw-LOC
 ‘Who whipped Mayaw?’ (in-situ Q=SUBJ.agt)
- b. [Mi-palo’-ay ci Panay [**cima-an**]_{OBL/FOC}]_{IP}?
 AV-whip-REAL PN Panay who-LOC
 ‘Who did Panay whip?’ (in-situ Q=OBL.pt)

(30) fronted Q (Central Amis - Formosan) (Yeh, fieldwork data)

- a. [U **cima**]_{FOCUS-C} [ku mi-palo'-ay — ci Mayaw-an]_{COMP}?
 PRT who CN.ABS AV-whip-REAL SUBJ PN Mayaw-LOC

(PIVOT)

‘WHO was the one that whipped Mayaw?’ (fronted Q=SUBJ/PIVOT.agt)

- b. [U **cima**]_{FOCUS-C} [ku ma-palo'-ay ni Panay —]_{COMP}?
 PRT who CN.ABS PV-whip-REAL PN.GEN Panay SUBJ

(PIVOT)

‘WHO was the one that Panay whipped?’ (fronted Q=SUBJ/PIVOT.pt)

- c. * U cima ku mi-palo'-ay ci Panay?
 d. * Cima-an ku mi-palo'-ay ci Panay?

Likewise, the difference between SUBJ and PIVOT is evidenced by the distinct status of SUBJ PIVOT in (30) and SUBJ in (29). In the latter, no FOCUS-C is involved, and thus SUBJ remains in-situ. By contrast, the fronted Qs in (30) specifically privilege SUBJ PIVOT as seen by the verbal voice morphology (e.g. unacceptability of 30c in contrast to 30b), and they are associated with the extra-syntactic function FOCUS-C in the discourse. Pragmatically, there is a difference between the fronted Qs and in-situ Qs. The in-situ Q in (29a) forms an open question without the presupposition of contrasting entities in the given context (Wei 2009: 348). In contrast, the fronted Q in (30a) is used when the SUBJ agent in question is one among a group of people present in a given situation. This indicates that the fronted Q comes with a pragmatic meaning of contrast that is not present with the in-situ Q.

Note that Amis differs from Balinese in that in-situ Qs in Balinese are OBJ, whilst those in Amis can be either SUBJ or non-SUBJ. This is because the two languages differ in their word order. SUBJ is pre-verbal in Balinese, whereas Amis is verb-initial like Squiliq and Tagalog (cf. Section 3) and thus, SUBJ is realised pre-verbally in pragmatically marked constructions.

In particular, the essence of PIVOT as the overlay function for clause-combining is evidenced by the structure of fronted Qs in Amis. Structurally, the sentences with fronted Qs in (30a–b) are pseudo-clefts in a bi-clausal structure. The Qs are fronted nominal predicates in FOCUS, followed by a headless relative clause flagged by *ku* (i.e. the ABS case nominal marker) in which SUBJ is obligatorily relativised. The SUBJ marker supplies the pronominal value that is coreferential with the fronted Q (cf. Section 6.3 for the LFG representation of bi-clausal structures with a nominal predicate).

However, the Balinese data point in (28b) also shows that the SUBJ/PIVOT constraint is not necessarily related to clause combining. This is expected as FOCUS-C (the critical element of PIVOT) is pragmatically driven for communicative purposes, applicable to a mono-clausal sentence.

Q fronting interacts with verbal voice morphology. In Amis, only the most prominent argument (i.e. SUBJ) takes part in this PIVOT function for fronting Qs. For instance, when understood as A, its selection as SUBJ is indicated by the same AV morphology for an in-situ SUBJ, as in (29a), and a fronted SUBJ, as in (30a) (i.e. SUBJ/PIVOT). However, when the Q *cima* bears the patient role, its fronting (i.e. linking Patient as the SUBJ in FOCUS-C) requires PV morphology as seen in (30b). Retaining AV morphology on the verb renders the structure with a fronted Q ungrammatical, as seen in (30c). Likewise, in contrast to (29b), the structure is ungrammatical when the fronted Q *cima* is OBL marked by *-an*, as in (30d).

In short, we have seen how PIVOT as a syntactic-pragmatic function combines the syntactic property of SUBJ and the FOCUS-C function in giving rise to the SUBJ/PIVOT constraint associated with Q fronting in Balinese and Amis. Other behavioural properties targeting SUBJ as PIVOT typically encountered in Philippine-type and Indonesian-type AN languages include control/raising and relativisation (see Arka 2003a: 11-26).

Recent research in Indonesian relativisation demonstrates strong evidence that PIVOT is not always SUBJ.²⁷ The distinction between SUBJ and PIVOT in Indonesian receives further empirical support by the fact that OBJ can also be PIVOT, as seen in relativisation in (31). However, this OBJ relativisation through gapping (i.e. OBJ PIVOT) is highly constrained. It is only possible in a specific construction when both SUBJ and OBJ are highly salient with the presence of certain contrastive adverbs, such as *hanya* ‘only’, where the SUBJ-only constraint that is typically imposed in complex clause formation in Standard Indonesian is not maintained. Thus, while the agent *kamu* ‘2SG’ is SUBJ in (31), as evidenced from the verbal AV morphology, it is not the PIVOT for relativisation. Readers are directed to Arka (2021) for a detailed discussion of these relativisation facts in Indonesian, and the puzzles they pose for analysis.

- (31) Standard Indonesian (Arka 2021: 196)
 [Gadis [yang [(barangkali) [hanya kamu bisa menaklukkan ____]_{CP}]_{CP}]_{NP}
 girl REL perhaps only 2SG can AV.conquer
 ‘the girl who perhaps only you can control’

²⁷This is evident in relativisation in familiar languages, like English, where non-SUBJ can be PIVOT (i.e. gapped in relativisation).

4.3 Non-SUBJ functions: OBJ and OBL

In LFG, there are three non-SUBJ functions: OBJ, OBJ_θ, and OBL. In this subsection, we explore their realisation in AN languages, and show that distinguishing these three non-SUBJ functions is useful in the transitivity analysis of the AV patient, and in the analysis of Indonesian-type applicatives. This is because LFG's modular design and conception of GFs as 'natural' classes allow us to not only distinguish OBJ from OBL at the level of syntactic f-/a-structure, but also to capture the gradient nature of the OBJ-OBL distinction in Prototype theory (cf. Taylor 2003) and a core index analysis (Arka 2017). We begin with a characterisation of OBJs.

On the basis of cross-linguistic GF classifications, and research on syntactic prominence and semantic role associations (Comrie 1989, Bresnan 2001), we define OBJ syntactically as a class of core complements that is prototypically and thematically unrestricted. The syntactic property of complementation distinguishes OBJ from SUBJ since SUBJ is not a complement, and the coreness property differentiates it from OBL since OBL is not a core argument. Defining OBJ as a class of GF in this way allows us to capture the varied characteristics of OBJ cross-linguistically, but also within the same language (cf. Dalrymple & Nikolaeva 2011). It also allows us to identify language-specific object-like patterns, which provide empirical grounds for identifying different kinds of OBJ: prototypical or primary OBJ (thematically unrestricted OBJ) and secondary non-prototypical OBJ (also thematically restricted, and otherwise known as OBJ_θ in LFG) (Bresnan & Kanerva 1989, Haspelmath 2007). In what follows, we show the variation in the actual morphosyntactic realisations of different types of OBJ in AN languages, starting with the prototypical OBJ.

The prototypical OBJ in descriptive/typological linguistics is patient-like in its semantic role. In our LFG analysis, this OBJ is linked to the a-object (i.e. <2:pt>) in the a-structure representation. In AN languages with voice systems, it is the core argument of the verb in the AV structure, and typically appears postverbally, like the NP *Watan* in (7) (Squiliq Atayal) and *apa* 'what' in (28) (Balinese). Squiliq Atayal and Balinese represent languages where free OBJ arguments have no specific OBJ flagging. OBJ NPs are bare, in contrast to prepositionally flagged OBLs.

However, there are also AN languages that specifically flag arguments with non-SUBJ core status, like *ng=* in Tagalog in (23) above, and *te* in *Tukang Besi* in (32) below. In *Tukang Besi*, the pronominal indexing system on the main (finite) verb of an embedded clause shows diminished voice morphology (Donohue 2008: 8). The underlying <2:pt> 'you' surfaces as OBJ in (32a) and is flagged by *te*, and not indexed on the verb. It appears as SUBJ, which is indexed by the enclitic *=ko*, and is optionally cross-referenced by the NOM NP that is flagged by *na* in (32b).

(32) Tukang Besi (WMP, Indonesia) (Donohue 2002: 85)

- a. No-kiki'i [te iko'o]_{OBJ} [na beka]_{SUBJ}.
 3REAL-bite CORE you NOM cat
 'The cat bit you.'
 SUBJ OBJ
 'bite<1:agt:'cat',2:pt:'you'>'
- b. No-kiki'i[=ko]_{SUBJ} ([na iko'o]_{SUBJ}) [te beka].
 3REAL-bite=2SG.OBJ NOM you CORE cat
 'The cat bit you.' or 'You, the cat bit.'
 OBJ SUBJ
 'bite<1:agt:'cat',2:pt:'you'>'

Note that the GF alternation in Tukang Besi in (32) is equivalent to the AV-UV alternation in Indonesian-type languages, like the Balinese example in (5). The key differences relate to verbal voice marking and argument flagging. Unlike in Balinese, the AV structure in Tukang Besi in (32a) has no verbal AV morphology, and its OBJ is overtly flagged.

The thematically unrestricted property of OBJ is captured by the $[-r]$ feature in LMT (Bresnan & Kanerva 1989; Dalrymple 2001: 21). That is, it is linkable to a range of roles other than patient. In our definition in this chapter, it is indeed a non-SUBJ core argument, as seen in Tukang Besi in (32b) where the OBJ flagged by *te* is linked to the agent. Additionally, other roles associated with OBJ include instrumental, benefactive/recipient, goal, and locative, as seen in the Indonesian-type languages that show applicative morphology (e.g. Indonesian, Balinese, Madurese, among others). Madurese has two applicative suffixes, namely *-e* (for locative/goal applicative) and *-agi* (for benefactive/instrumental), both of which are equivalent to *-i/-kan* in Indonesian (Arka et al. 2009) and *-in/-ang* in Balinese (Arka 2003a). The Madurese examples in (33) show that the post-verbal OBJ is the thematically unrestricted OBJ, which is linked to patient/theme in (33a), locative/goal in (33b) (with the verb containing the locative applicative, *-e*), and recipient/benefactive in (33c) (with the verb containing the recipient applicative, *-agi*).

(33) Madurese (Davies 2010: 283, 299)

- a. Embuk ngerem [paket]_{OBJ} [ka Ebu']_{OBL}.
 elder.sister AV.send package to mother
 'Big Sister sent a package to Mother.'
 SUBJ OBJ OBL
 'AV.send<1:agt,2:pt |3:goal>'

- b. Embuk ngerem-e [Ebu']_{OBJ} [paket]_{OBJ_θ}.
 elder.sister AV.send-APPL mother package
 'Big Sister sent Mother a package.'
 SUBJ OBJ OBJ_{THEME}
 'AV.send.for<1:agt,2:goal, 3:th>'
- c. Sa'diyah melle-yagi [na'-kana']_{OBJ} [permen]_{OBJ_θ}.
 Sa'diyah AV.buy-APPL REDUP-child candy
 'Sa'diyah bought the children candy.'
 SUBJ OBJ OBJ_{THEME}
 'AV.send.for<1:agt,2:goal, 3:th>'

While primary OBJ is thematically unrestricted, secondary non-prototypical OBJ is typically thematically restricted. This is evidenced in the important effect of applicativisation whereby the a-structure is restructured with OBJ and OBJ_θ surfacing differently. Consider, firstly, the PP in (33a), *ka Ebu* 'mother', which is prepositionally flagged as OBL (i.e. non-core). Yet, the argument is promoted to the secondmost prominent slot in the applicative structure in (33b). Its realisation as a bare NP, and its structural position immediately following the verb, indicate that it is OBJ, while the underlying displaced theme *paket* is demoted to the third core position, and surfaces as OBJ_θ. This results in a ditransitive structure of SVOO. Likewise, the same restructuring of a-structure occurs with the benefactive applicative in (33c).

In LFG, the NP *paket* in (33b) is an instance of OBJ_{THEME} in Madurese. Semantically, it is restricted to a displaced theme only. Crucially, and unlike OBJ (*Ebu*), it is restricted in the sense that it does not surface as SUBJ in the UV voice, as seen in the ungrammaticality of (34b) in contrast to (34a). This provides clear evidence that the applied argument occupies the second argument in the restructured transitive a-structure. Hence, it is 'mappable' to OBJ in AV in (33b), or SUBJ in uv in (34a).²⁸

(34) Madurese (Davies 2010: 284)

- a. Ebu' e-kerem-e [paket]_{OBJ} bi' Embuk.
 mother UV-send-APPL package by elder.sister
 'Mother was sent a package by Big Sister.'
 OBJ SUBJ OBJ_{THEME}
 'uv.send<1:agt,2:goal, 3:th>'

²⁸The preposition *bi* 'by' is optional in Madurese. There is no identifiable grammatical difference between the pairs with/without *bi*; the verb in this structure is therefore analysed as UV, not passive (Davies 2010: 256-258).

- b. *[Paket rowa]_{SUBJ} e-kerem-e (ka) Ebu' bi' Embuk.
 package that UV-send-APPL to mother by elder.sister
 ('The package was sent (to) Mother by Big Sister.')

In AN languages with a systematic argument indexing system, OBJ is typically semantically transparent from its case form. That is, the OBJ index is part of a verbal complex structure, either as a pronominal affix or clitic, and surfaces differently according to semantic roles. In Kambera, for example, the prototypical patient-like OBJ is expressed by an ACC enclitic immediately following the verb, whereas the benefactive OBJ is marked differently via DAT. Hence, the first-person patient OBJ is *ka* '1SG.ACC' in (35a), but *ngga* in (35b), since it is thematically beneficiary. Note that the displaced theme, OBJ_{THEME}, in (35b) is DAT. In LFG, the Kambera ditransitive sentence in (35b) has the same a-/f-structures as the Madurese examples in (33b–c), with the key differences being in the coding and feature values of the surface GFS. For the right enclitic form of OBJ to be selected, the lexical entry must be specified by the relevant constraints, as shown in (35c) with =*ngga*. The shorthand $(\uparrow \text{OBJ})_{\sigma} = (\uparrow_{\sigma} 2:\text{ben})$ constraint relies on a sigma projection relating f-structure to a-structure, here establishing a correspondence between the OBJ in the f-structure and the second benefactive argument in the a-structure (see Belyaev 2023b [this volume] for discussion of LFG's projection architecture).

(35) Kambera (Klamer 1998: 63)

- a. (Na tau wútu) na=palu=ka (nyungga).
 ART person be.fat 3SG.NOM=hit=1SG.ACC I
 'The big man hit me.'
 SUBJ:nom OBJ:acc
 'hit< 1:agt, 2:pt>'
- b. (I Ama) na=kei=ngga=nya.
 ART father 3SG.NOM=buy=1SG.DAT=3SG.DAT
 'Father buys it for me.'
 SUBJ:nom OBJ:dat OBJ_{THEME}:dat
 'buy< 1:agt, 2:pt 3:th>'
- c. *ngga* CLITIC $(\uparrow \text{OBJ PRED}) = \text{'PRO'}$
 $(\uparrow \text{OBJ PERS}) = 1$
 $(\uparrow \text{OBJ NUM}) = \text{SG}$
 $(\uparrow \text{OBJ CASE}) = \text{DAT}$
 $(\uparrow \text{OBJ})_{\sigma} = (\uparrow_{\sigma} 2:\text{BEN})$

The different coding of OBJ, as seen in Kambera, is not typologically unusual. It is known as Differential Object Marking (henceforth DOM) (Dalrymple & Nikolaeva 2011). For instance, Palauan has DOM that is primarily regulated by semantic features. However, unlike Kambera, Palauan demonstrates DOM that is determined by definiteness, instead of semantic roles. Definite OBJ receives pronominal indexing on the verb, as in (36a), whereas indefinite OBJ does not, as in (36b). In an LFG analysis, Palauan DOM can be captured by annotating the suffix slot in the verb formation rule with the constraining equation: $(\uparrow \text{OBJ DEF}) = c+$. The suffix *-ii* also carries a definiteness feature in its lexical entry, $(\uparrow \text{DEF}) = +$, in addition to person and number features.

(36) Palauan (WMP, Palau) (Georgopoulos 1991: 45)

- a. Te-'illebed-ii a bilis a rengalek.
 3PL-PFV.hit-3SG dog children
 'The kids hit the dog.'
- b. Te-'illebed a bilis a rengalek.
 3PL-PFV.hit dog children
 'The kids hit a dog/the dogs/some dog(s).'

Oblives in AN languages are typically phrasally flagged. The common pattern is that OBL is flagged by an adposition, like *ka* 'to' for OBL locative/goal in Madurese (33a), and *teken* 'by' for OBL agent in Balinese (Arka 2019: 262). However, AN languages of the Philippine type have phrasal markers that specifically mark OBL status in contrast to the core status of SUBJ. This is the case in Puyuma where the OBL and SUBJ are equally flagged. However, Puyuma shows differential OBL marking on the basis of differences in nominal type (e.g. common versus proper) and definiteness (as seen in DOM) rather than differences in semantic roles. Consider example (37) below, where *kana* is used as the phrasal OBL marker for a definite common noun like in (37a), and *dra* for an indefinite common noun as in (37b–d). The same phrase marker, *dra*, is used for indefinite obliques irrespective of their roles as patient, instrument, location, etc.

(37) Puyuma

- a. Ku=tuLud-anay na sarekuDan kana temumuwan.
 1SG.GEN-pass-IV DEF.NOM stick DEF.OBL offspring
 'I passed the stick to the offspring.' (Teng 2005: 23)
- b. Traka-trakaw=ku dra akan-an.
 <AV>REDUP-steal=1SG.NOM INDF.OBL eat-NMLZ
 'I stole food repeatedly.' (Teng 2008: 146)

- c. Tu=pa-ladram-aw **dra** lrangetri pa-karun.
 3GEN=CAUS-know-PV INDF.OBL stick CAUS-work
 ‘They used a stick to teach them to work.’ (Teng 2008: 245)
 (translation adapted)
- d. Ka-sa-sanan **dra** dalran.
 ka-REDUP-stray INDF.OBL road
 ‘He will get lost.’ (Teng 2008: 168)

4.4 Alignment systems and related phenomena

The syntactic status of the non-SUBJ argument is relevant to the question of alignment. There is a long-standing debate in AN linguistics as to whether syntactic alignment has properties of ergativity, accusativity or split-ergativity. There are competing proposals in the literature, as well as claims that Western AN languages vary in their alignment; see Aldridge (2004), Katagiri (2005), and references therein for further discussion. In the following section, we present cases where morphosyntactic ergativity is firmly observed, like in Puyuma, then move to borderline cases.

Puyuma exhibits syntactic properties that are typical for an ergative system. However, unlike well-known ergative languages such as Dyirbal (Dixon 1972), there are no morphologically ‘basic’ or unmarked transitive verbs in Puyuma because they are all marked for their specific non-actor voices; e.g. *-anay* marking for CV, conveyance voice, in (37a), and *-aw* for PV, patient voice, in (37c). The AV verbs are also morphologically marked by *-em-* as in (37b).²⁹ The AV structure can be analysed as antipassive because the patient argument of the transitive verb is demoted to non-core status, which is flagged by the OBL marker as shown in (38). Puyuma, therefore, exhibits clear syntactic asymmetry in its voice alternations, which is the hallmark of a truly ergative system. In the transitive structure, <1:agent, 2:patient/theme>, the two core arguments are not equally selectable as syntactic SUBJ/PIVOT. That is, SUBJ/PIVOT selection is asymmetrically aligned towards the second patient core slot. Hence, when the agent has to be linked to SUBJ/PIVOT, the patient must be removed and demoted to non-core status in order to allow for the linking of the agent to SUBJ/PIVOT. Removing the patient from the core status in the a-structure results in an intransitive <1:agent | 2:patient/theme> structure.

²⁹Note that in Teng’s (2005, 2008) descriptions, the AV affix *-em-* is glossed as intransitive (INTR) because the AV structure is syntactically intransitive.

(38) Puyuma (Teng 2008: 72, 187)

- a. T⟨em⟩engedr=ta dra unan i, ...
 ⟨AV⟩kill=1PL.NOM INDF.OBL snake TOPIC, ...
 ‘We killed a snake, ...’
- b. K⟨em⟩asu=ta dra eraw, dra irupan.
 ⟨AV⟩bring=1PL.NOM INDF.OBL wine INDF.OBL dishes
 ‘We brought some wine and some dishes.’

However, in many other AN languages with robust voice morphology, the antipassive analysis of AV is controversial because the evidence for the demotion of the underlying patient to OBL is often unclear and debatable. In Tagalog, for example, the patient argument of the AV sentence is flagged by the core phrase marker *ng* in (23a). Thus, the AV sentence in Tagalog is distinct from Puyuma in that it is syntactically transitive. Conversely, under an ergative analysis, the AV is analyzed as antipassive on the basis that P is understood as indefinite, which is a typical semantic property of the antipassive patient (Hopper & Thompson 1980). Yet, this semantic criterion for the core status of AV patient is disputable, as shown in the Paiwan (Formosan) examples in (39) below. While an OBL patient may be indefinite, as in (39a), the reverse does not hold since an oblique-marked patient can have a definite reading, as seen in (39b) (cf. DOM in Puyuma in Section 4.3). This suggests that in many Philippine-type languages, the core-ness status of the non-SUBJ argument in AV cannot be easily and solely specified by its semantic property due to the mismatch of semantic transitivity, syntactic transitivity and voice alternation.³⁰

(39) North Paiwan (Formosan) (Chang 2006: 114, 412)

- a. Ki-lakarav tua sipangetjez tua zua marekaka.
 obtain.AV-flower OBL.CN gift OBL.CN that both.sibling
 ‘(He) would pluck flowers as a gift for both sisters.’
- b. Na=t⟨em⟩ekeL=anga timadju tua ?ucia.
 PFV=drink⟨AV⟩=COMPL 3SG.NOM OBL.CN tea
 ‘He has drunk the tea.’

³⁰The status of coreness must, therefore, be determined by taking into account all the relevant language-specific morphosyntactic properties. This is possible via a core index analysis (Arka 2017), for example. The core index analysis applied to the P of the AV structure in Puyuma reveals a core index of 0.44, which is classified as OBL albeit atypical. A prototypical OBL in Puyuma (e.g. LOC OBL of the AV verb) has a core index of 0.11, which is in line with the cross-linguistic tendency for prototypical OBL to have a core index of below 0.20. The degrees of coreness/obliqueness for the P of AV structures across other Philippine-type languages is a matter of future research.

Likewise, for AN languages in the regions of Sulawesi, which have been analysed as showing ergative properties, the status of the AV patient is not very straightforward either. Consider the examples in (40) from Moronene (in Southeast Sulawesi). Moronene shows DOM whereby a definite OBJ NP receives object indexing. Conversely, an indefinite or a non-specific OBJ NP receives no such indexing. The AV sentence with AV morphology (*moN-*) has been analysed as antipassive (Andersen & Andersen 2005) based on the patient NP being indefinite or non-specific.

(40) Moronene (WMP, Indonesia) (Andersen & Andersen 2005: 246, 252)

- a. Yo laku ari **kea'-o** manu.
ART civet already bite-3SG.ABS chicken
'The civet bit the chickens.' [laku11]
- b. Da-hoo nta **mong-kea** miano.
be-3SG.ABS FUT AV.NF-bite person
'It will bite someone.' [col85] [AuAbmV]

While it is true that the AV structure shows a lower degree of transitivity in terms of parameters described by Hopper & Thompson (1980), it is not syntactically antipassive in the analysis where the a-structure consists of two core arguments; that is, the patient NP in (40) is OBJ, not OBL. Additional evidence for this comes from its expression in bare NPs and the fact that OBL is prepositionally flagged in Moronene.

In addition to semantic properties, other syntactic evidence in complex constructions such as control properties has been used to argue for an accusative and/or a split ergativity analysis (i.e. ACC case for the AV patient). For instance, proponents of treating AV patient as a core argument (Hsin 1996, Chang 2000) would analyse the phrasal marker *tu* in Kavalan (41) as an accusative case marker, as it phrasally marks the a-object of *pumupup* that functionally controls the subject of the second verb *matiw* 'go'. The argumentation here is that only core status can allow an argument to be the controller.

(41) Kavalan (Chang 1997: 198)

- P<um>upup tina-na **tu sunis 'nay** m-atiw sa Bakung.
<AV>persuade mother-3SG.GEN OBL child that AV-go PREP Bakung
'That child's mother persuaded the child to go to Bakung.'

However, the status of controller in the matrix clause may not be decided purely on syntactic grounds since it also depends on the semantic properties of the matrix verb. The control construction in (41) is analysed as the 'influential' type

of control, defined by lexical semantic properties (Sag & Pollard 1991), where the controller is the influenced argument (i.e. the persuadee) regardless of its GF. In other terms, the choice of controller is based on the lexical semantics of the control verb that requires an intentional agent in an open clause complement (xCOMP). However, the control verb does not specify that the syntactic properties of the influenced argument are core or oblique.

Instead, it is the status of the controllee in the embedded clause that provides the diagnosis of termhood (Kroeger 1993: 40). Only core arguments can be the controllee, as opposed to OBL arguments. In the Kavalan example (41), the controllee is the SUBJ of a non-finite AV verb *matiw*, so the agent argument in the embedded clause fulfils both syntactic and semantic properties required by the 'influential' type of control. Likewise, in Haian Amis, the core status of the AV agent is evident by its property as the controllee (i.e. AV-SUBJ) as in (42a–b), regardless of its status as the controller (i.e. PV-SUBJ or AV-OBL) in the matrix clause. By contrast, as shown in (42c), it is not acceptable for the AV patient (i.e. *ci Akian* in 42c) in the embedded clause to be the intended controllee.³¹

(42) Haian Amis (Formosan) (Wu 2006: 378–379)

- a. Ma-ucur aku ci Aki mi-to'or ci Panay-an.
PV-assign 1SG.GEN PN.ABS **Aki** AV-follow PN Panay-LOC
'I assigned Aki to follow Panay.'
- b. Mi-ucur kaku ci **Aki-an** mi-to'or ci Panay-an
AV-assign 1SG.ABS PN **Aki-LOC** AV-follow PN Panay-LOC
'I am going to assign Aki to follow Panay.'
- c. *Mi-ucur kaku ci **Aki-an** mi-to'or ci Panay
AV-assign 1SG.ABS PN **Aki-LOC** AV-follow PN.ABS Panay
'I am going to assign Aki to be followed by Panay.'

In comparison to other grammatical tests (e.g. 2P clitic placement, pronominal bound forms and DOM), evidence of control in complex constructions for testing the status of non-SUBJ arguments should be examined carefully. That is, using control verbs as the evidence of an accusative analysis of the AV construction necessitates a meticulous evaluation and differentiation between properties of control that have semantic roots and those that are purely syntactic in nature.

At the morphological level, pronominal forms (affixes/clitics) across AN languages show nominative and ergative alignment. In AN languages with robust

³¹However, Haian Amis differs from Tagalog (Kroeger 1993) and Pazeh-Kaxabu (Yeh in preparation) in that the core status of a PV agent cannot be observed via properties of the controllee.

voice systems, the bound pronouns typically consist of two sets. The first set is often labelled GEN, or ERG under an ergative analysis. This pronominal form is linked to the transitive agent argument in non-AV structures, such as *ku=* and *tu=* in Puyuma in (37), *=na* in Kavalan in (6), and *no-* in *Tukang Besi* in (32). The second set is the SUBJ/PIVOT form and is typically labelled NOM in the AN literature. This is the thematically unrestricted form that is linkable to any semantic role of a core argument, including the patient core argument of a transitive verb and the intransitive subject. This justifies the labelling of this set as the ABS(olutive) form in an ergative analysis, as exemplified by *-(ho)o* in Moronene in example (40) above. Note that in descriptive works, such as Teng's (2008) description of Puyuma, the second set is also (confusingly) called the NOM(inative) set even though the language shows an ergative alignment property. Morphology and syntax in LFG are separate modules in grammar with case (marking) being dealt with at the morphology-syntax interface (see Bresnan & Mchombo 1987). It is captured through the CASE feature constraint, which is associated with GF linking. Thus, in a language like Puyuma and Pazeh-Kaxabu where there is empirical evidence for ergative alignment (both morphologically and syntactically), a pronominal affix/clitic can be specified as having a CASE feature in its entry: (\uparrow CASE)=ABS. The grammar of the language can be globally specified as having a conditional if-then constraint: (\uparrow SUBJ) \Rightarrow (\uparrow SUBJ CASE)=ABS. Because this constraint applies to verbs broadly, one way to handle it is by incorporating it into the rule that introduces the clausal c-structure that comes with the SUBJ annotation. This constraint means that if the argument is selected as SUBJ then it must have ABS case. Other pronominal clitics can be specified as having (\uparrow CASE)=ERG in their entries for languages like Puyuma, and specifically for the agent a-subject argument of a transitive predicate. However, for other languages that show a SUBJ fixed linking with NOM-ACC alignment, as in Kambera (Klamer 1998: 73), a different specification must be given for the pronominal clitic linked to the transitive agent argument, namely (\uparrow CASE)=NOM.

For non-pronominal forms, the semantic and syntactic information throughout the system can be specified in the entry for phrasal markers.³² This applies to the differential OBJ and OBL marking, noting that we extend DOM to include OBL marking as well). For simplicity, only one marker of DOM is exemplified below

³²The term 'phrasal markers' finds frequent usage in AN linguistics, especially when characterizing Philippine-type AN languages. These markers, like *na* and *kana*, which mark SUBJ and definite OBL relations in Puyuma (as seen in example (37a)), tend to manifest in diverse forms across various AN languages. They are often labelled differently by different authors depending on their analysis, such as clitics, case markers, non-/personal markers, or prepositions (Himmelman 2005: 144–149).

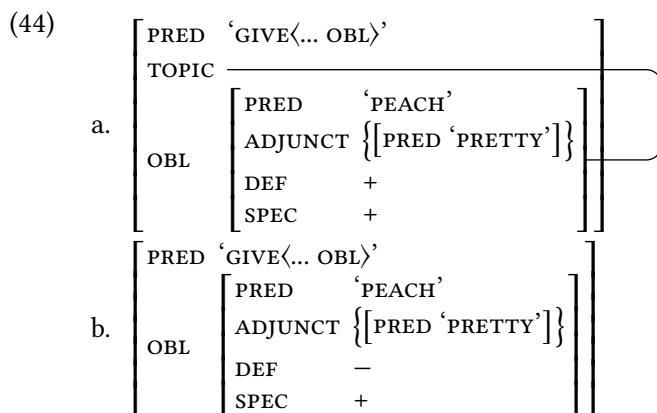
in (43). DOM across languages commonly draws on different semantic properties. In LFG, these semantic features can be specified together with the semantic case value without affecting the syntactic status of the argument (Butt & King 1991, 2003, Dalrymple & Nikolaeva 2011).

In Pazeh-Kaxabu, DOM encodes the differential information related to the topicality and specificity/definiteness of P (Yeh in preparation). For example, in the AV structure in (43a), P is realised as OBL due to the ergative system and is flagged by *u* because the referent is definite and topical (reading [i]) or specific indefinite (reading [ii]). The non-specific indefinite P (cf. example 10b) is realised by an unmarked bare NP.

Extending from the classic, integrated *i*- and *f*-structure (cf. Bresnan & Mchombo 1987) we represent the simplified lexical entry of the phrase marker *u* in (43b). It specifies a constraint that the noun phrase flagged by *u* must be OBL whose CASE is LOC. In addition, it imposes a disjunctive specification with two options capturing the two readings in (43a). The first option in reading (i) reflects sharing of the values of OBL argument and TOPIC. This is shown in the partial *f*-structure in (44a) where the reference is definite and specific (cf. Enç 1991; see also von Stechow 2002 for the distinction and interaction of definiteness and specificity). The (partial) *f*-structure for reading (ii) is given in (44). It captures the crucial difference in that there is no sharing as the OBL is indefinite and not TOPIC. The empirical fact about having in-/definite readings in the OBL argument (cf. also the Paiwan example in (39) above) is elegantly shown without assuming syntactic status to be determined by semantic property.

(43) Pazeh-Kaxabu (Li & Tsuchida 2002: 169)

- a. ... babaxa u kia'aren a arim.
 AV.give LOC pretty LNK peach
 (i) '... gave the pretty peach(es).' or
 (ii) '... gave certain pretty peaches.'
- b. *u* (OBL ↑)
 (↑ CASE)=LOC
 {(TOPIC ↑) (↑ DEF)=+ (↑ SPEC)=+
 | (↑ DEF)=– (↑ SPEC)=+ }



There are AN languages showing properties of split intransitivity or split-S. The split can be reflected in the argument pronominal marking, as in Acehnese (Durie 1987), or the morphological marking on verbs, which correlates with the properties of semantic roles as well as lexical-aspectual properties. Acehnese, for example, is an AN language with systematic clitic sets that cross-reference A(ctor) versus U(ndergoer) roles (Durie 1987). It has a split/fluid S or active system, as seen in examples (45)-(46) below. SUBJ in Achenese is, therefore, semantically very transparent and not a neutralised or syntactic SUBJ/PIVOT as seen in Philippine/Indonesian types. It is not uniquely picked up by a set of morphosyntactic behavioural properties, such as 'control' (see Section 5.1). LFG is well-equipped to handle such kinds of split transitivity (cf. Zaenen 1993; Arka 2003a). For example, the A and U clitics must have a linking constraint specified in their lexical entries, as shown in (45c) and (46c), respectively. The sigma metavariable (\uparrow_σ) in the entries ensures the correct mapping or correspondence between semantic a-structure and f-structure, so the constraint represented as (\uparrow_σ A) in (45c) states that semantically *geu* must be Actor). In addition, the specification (\uparrow SUBJ) for *geu* also ensures that it is associated with SUBJ. However, the undergoer or P clitic, *geuh*, must have a disjunctive specification to capture the fact that a sole argument (S, or SUBJ) of an intransitive verb has the same form as the undergoer (P, or OBJ) in a transitive clause (i.e. S_p/P pattern of the split).

(45) Cross-reference Actor (Acehnese, WMP, Indonesia) (Durie 1987: 366)

- a. Gopnyan **geu**=mat lôn.
 3 3A=hold 1SG
 'S/he holds me.'

- b. **Geu**=jak gopnyan.
 3A=go 3
 'S/he goes.'
- c. *geu* CL (↑ PRED)= 'PRO'
 (↑ PERS)=3
 (↑ NUM)=SG
 (↑ SUBJ)_σ = (↑_σ A)

(46) Cross-reference Undergoer (Acehnese) (Durie 1987: 369)

- a. Gopnyan ka lôn=ngieng=**(geuh)**.
 3 IN 1SG.A=see(=3P)
 'I saw him/her.'
- b. Gopnyan rhêt(=**geuh**).
 3 fall(=3)
 'S/he falls.'
- c. *geuh* CL (↑ PRED)= 'PRO'
 (↑ PERS) = 3
 (↑ NUM) = SG
 (↑ {SUBJ|OBJ})_σ = (↑_σ P)

This section has demonstrated how AN languages differ in their development of the voice system, and how they also show variation in the realisation of grammatical functions and DOM patterns. The theoretical advances in LFG studies—such as the inventory of GFS, the syntacticised a-structure, the overlay function PIVOT, and the specifications of case, information status and referential semantics—have shown advantages in capturing some patterns in AN languages that have long been controversial.

5 Complex constructions

Following the discussion of word order and the basic notions of how AN morphosyntax is represented in LFG, we now move on to some complex constructions. In this section, we highlight two salient features of complex structures in AN languages which are of long-standing theoretical and typological interest: complementation that involves argument gapping or control in the embedded clause, and complex predication with a particular focus on SVCs.

5.1 Complementation and control

Complement clauses are object-like clausal arguments which, for certain matrix verbs, may be syntactically peripheral or oblique-like. Formally, and in LFG terms, they are realised as COMPS (finite clauses) and xCOMPS (non-finite clauses with syntactic SUBJ-control). The distinction between COMP and xCOMP and their core status may not always be easy to identify. In what follows, we outline clear cases of (x)COMPS and their syntactic status.

Languages with robust voice morphology provide a diagnostic tool to determine the core status of (x)COMP. For example, in Indonesian-type languages, only a core argument can be selected as SUBJ/PIVOT, and a peripheral oblique/adjunct-like argument must be promoted to become a core argument in order to be realized as SUBJ. This is the case with the Balinese verb *edot* ‘want’. It is a two-place intransitive verb with the second argument being either a simple oblique argument appearing as a PP, like in (47a) below, or an xCOMP (without P-flagging) as in (47b). In both cases, the applicative *-ang* cannot be used. However, when an embedded clause is fronted and given the discourse function FOCUS-C (i.e. made the PIVOT/SUBJ), as in (47c), the applicative *-ang* is obligatory; the verb *edot=a* is unacceptable. Note, however, that the matrix verb must be in UV since the AV form *ng-edot-ang* ‘AV-want-APPL’ is unacceptable. That is, the clausal argument is treated as a non-Actor core argument. The obligatory applicativisation serves as evidence that the second clausal (COMP) argument with *edot* is syntactically oblique-like in (47b), but a core argument in (47c).

(47) Balinese (Arka 2003b: 135)

- a. Ia edot / *edot-ang [teken poh]_{OBL}.
 3 want want-APPL to mango
 ‘(S)he wants a mango.’
- b. Ia edot / ?*edot-ang [___ ngae umah lung]_{xCOMP}.
 3 want want-APPL SUBJ AV.build house good
 ‘(S)he wants build a good house.’
- c. [___ Ngae umah lung]_{PIVOT} (ane)
 SUBJ AV.build house good FOC
 edot-ang=a/*edot=a/*ng-edot-ang.
 UV.want-APPL=3
 ‘Building a good house is what s/he wants.’

However, xCOMP can also be a core argument. This is the case with the xCOMP of the verb *coba* ‘try’ in Indonesian in (48a) below. In Indonesian, like in Balinese,

an OBL cannot alternate with SUBJ/PIVOT without applicativisation. The xCOMP of the verb *coba* can, however, alternate to become SUBJ/PIVOT without applicativisation, as seen in (48b). Note that *coba* ‘try’ allows different patterns of control, including the so-called double (backward/forward) control given in (48c). This double control structure shows two gaps—left-headed and right-headed arrows indicate backward and forward control types, respectively (see Arka 2000 and Arka 2014a for details).

(48) Indonesian (Arka 2014a: 31 and own knowledge)

- a. Aku sudah mencoba [___ menjual mobil itu]_{xCOMP}.
 1SG PFV AV.try AV.sell car that
 ‘I have tried to sell the car.’
- b. [___ Menjual mobil itu]_{PIVOT} yang sudah ku=coba.
 AV.sell car that FOC PFV 1SG=UV.try
 ‘Selling the car is what I have tried (to do).’
- c. [Mobil itu]_{PIVOT} (yang) sudah ___ coba [___ ku=jual]_{xCOMP}.
 car that FOC PFV (A) UV.try (P) 1SG=UV.sell
 ‘That car (is the one that) I have tried to sell.’

The clausal argument of a raising verb can also have an xCOMP with the raised argument being obligatorily SUBJ. The following shows the (unusual) SUBJ raising to matrix OBL in Puyuma. In (49a) below, the clausal complement of the verb ‘know’ is COMP. It is syntactically non-core since it is flagged by *dra* (i.e. the indefinite OBL phrase marker; glossed as a complementiser for clarity here). The patient NP ‘the fish’ (indicated in bold) is selected by the PV *-aw* on the verb as SUBJ, and present in the embedded clause. In (49b), however, the SUBJ is raised and appears as OBL in the matrix clause, flagged by *kana*. Note that raising in (49b) is not possible with an embedded verb containing the voice suffix, *-anay*, since this selects an instrumental argument instead.

(49) Puyuma (Teng 2008: 153-154)

- a. Ma-ladram=ku [dra tu=lriputr-aw na **kuraw**
 INTR-know=1SG.NOM COMP 3GEN=wrap-PV DEF.NOM fish
 dra bira’]_{COMP}.
 INDF.OBL leaf
 ‘I know that the fish was wrapped in a leaf.’

- b. Ma-ladram=kun kana kuraw [dra tu=lriputr-aw/
 INTR-know=1SG.NOM DEF.OBL fish COMP 3GEN=wrap-PV/
 *tu=lriputr-anay dra bira']_{XCOMP}
 3GEN=wrap-IV INDF.OBL leaf
 'I know that the fish was wrapped in a leaf.'

In the indexing AN languages of eastern Indonesia and Oceania, the SUBJ (bound) pronoun is typically part of the verbal morphology and cannot be gapped. SUBJ is not a syntactic PIVOT for clause combining purposes in these languages. There is, therefore, no syntactic control or raising. Clausal arguments are consistently COMPS with no XCOMP alternative. This is the case in Taba in (50), and Mangap-Mbula in (51):

- (50) Taba (CEMP, Eastern Indonesia) (Bowden 2001: 391)
 Nculak wangsia de lmul akle.
 n=sul-ak wang=si de l=mul ak-le
 3SG=order-APPL child=PL RES(so.that) 3PL=return ALL-land
 'He told the children to go home.'
- (51) Mangap-Mbula (Oceanic) (Bugenhagen 1995: 272)
 Ti-manjmanj yo [be anj-kam pizin].
 SUBJ:3PL-urge OBJ:1SG COMP SUBJ:1SG-do DAT.3PL
 'They urged me to give it to them.'

Despite their rarity, some Oceanic languages with indexing systems, such as Hoava (Davis 2003), Longgu (Hill 2002), and Kokota (Palmer 1999), have syntactic SUBJ/PIVOT. In Hoava, the index on the verb is only for OBJ. This language shows COMP, as in (52a), as well as XCOMP like in (52b). Complement-taking predicates in Hoava come with an invariant OBJ index *-a* which signals that there is an embedded complement clause in the structure.

- (52) Hoava (Oceanic) (Davis 2003: 288)
- a. Hiva-ni-a ria [de pule mae sa qeto]_{COMP}.
 want-APPL-OBJ:3SG 3PL COMP return come ART:SG war.party
 'They wanted the war party to come back.'
- b. Haku=haku-ni-a ria [de naqali-a]_{XCOMP}.
 REDUP=be.tired.of-APPL-OBJ:3SG 3PL COMP carry.TR-OBJ:3SG
 'They were tired of carrying it.'

5.2 Serial Verb Constructions

SVCs are the hallmarks of AN languages of the isolating type and are observed in the languages of eastern Indonesia, such as Rongga (Arka 2016), and also in Oceanic languages, as discussed below. Some SVCs are also encountered in the agglutinating Philippine/Indonesian-type languages, including Balinese (Indrawati 2014) and Puyuma (Teng 2008).

Unlike complementation, SVCs syntactically express complex (sub)events in monoclausal structures (Crowley 2002, Haspelmath 2016). Semantically, the relations between subevents typically convey adverbial modification with meanings such as comitative, benefactive and instrumental. However, they may also express other tightly integrated meanings often discussed under the rubric of complex predicates (see Arka & Simpson 2008). For example, the SVC expresses the desiderative ‘want’ (i.e. ‘feel-say’) in Ambae in (53), and the causative and resultative meaning in Rongga in (54) and Mwotlap in (55).

- (53) Ambae (Oceanic) (Hyslop 2001: 387)
 No=mo rongo vo na=ni qalo.
 SUBJ:1SG=REAL feel say SUBJ:1SG=IRR fight
 ‘I want to fight.’
- (54) Rongga (Arka 2016: 227)
 Selu tau mata manu ndau.
 Selu make die chicken that
 ‘Selus killed the chicken.’
- (55) Mwotlap (Oceanic) (François 2006: 232)
 Ne-lên mi-yip hal-yak na-kat.
 ART-wind PRF-blow fly-away ART-cards
 ‘The wind blew the cards away.’

SVCs can be analysed in LFG in the same way as complex predicates through predicate composition (Andrews & Manning 1999). The exact c-structure varies according to the language considered, but it is typically a compound-like nested structure: $[V(P)_1 V(P)_2]_{V(P)}$. That is, there is a higher VP consisting of lower VPs in the c-structure. The crucial idea of the analysis is to capture the empirical fact that the SVC is monoclausal; that is, the V(P) component(s) share the same SUBJ, and possibly another argument, depending on the transitivity of V_1 and V_2 verb components.

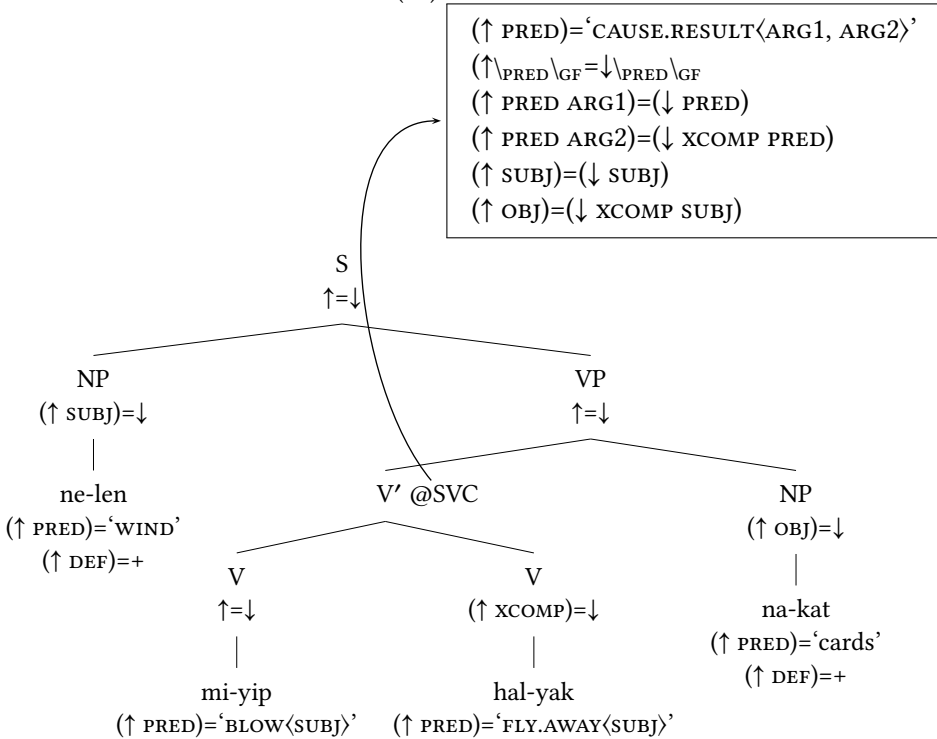
SVCs also reveal an intriguing and important property of the construction, exemplified by the Mwotlap example in (55). The causative-resultative meaning is constructed at the level of SVC because neither V_1 nor V_2 carry a causative-resultative meaning lexically. That is, the syntactic transitivity is constructional because neither V_1 nor V_2 is transitive. In LFG, such resultative constructions, as in example (55), can be captured by lexical-constructional a-structure, indicated by the SVC template, @SVC, annotated to V' of the VP in (56b). The template consists of complex equations given in the box showing the constructional predicate of 'CAUSE.RESULT<ARG1, ARG2>' (where ARG1 is the causing event and ARG2 is the resulting event). The restriction operator expressed by $\uparrow \backslash_{\text{PRED} \backslash_{\text{GF}}} = \downarrow \backslash_{\text{PRED} \backslash_{\text{GF}}}$ (Kaplan & Wedekind 1993) regulates the predicate composition involved in the SVC; see Butt et al. (2003) for the application of the restriction operator in Urdu/Hindi and other languages. This restriction operator and the other constraints associated with @SVC result in an f-structure with the subcategorisation frame shown in (56a).

Note that, in the resultative SVC of (56b), the OBJ annotation is specified at the NP of the higher VP, sister of V' , of the c-structure since it is the OBJ argument of the constructed causative-resultative predicate; neither V_1 nor V_2 has OBJ. The @SVC template (with detailed specifications provided in the box) specifies that the SVC's OBJ has the same value as the lower V_2 's SUBJ, and the SVC's SUBJ has the same value as the SUBJ of V_1 . This sharing of values for SUBJ and OBJ is indicated through tags [1] and [2] in (56a).

(56) a. f-structure of sentence (55)

$$\left[\begin{array}{l} \text{PRED} \text{ 'CAUSE.RESULT<SUBJ[1]: 'BLOW<[1]>', OBJ[2]: 'FLY.AWAY<[2]>' } \\ \text{SUBJ} \left[\begin{array}{l} \text{PRED 'WIND'} \\ \text{DEF +} \end{array} \right] \\ \text{OBJ} \left[\begin{array}{l} \text{PRED 'CARDS'} \\ \text{DEF +} \end{array} \right] \end{array} \right]$$

b. c-structure of sentence (55)



However, the distinction between mono-clausal SVCs and bi-clausal subordination is not always clear. A typical diagnostic test for SVCs is negation: since SVCs are monoclausal, the criterion of single negatability applies (Durie 1997). There are also other language-specific criteria that distinguish SVCs from multi-verb constructions in coordinate and subordinate clauses. In Balinese, for example, the presence/absence of voice morphology serves as a diagnostic criterion. The second verb in an SVC may optionally contain an AV prefix, indicated by putting the AV prefix in brackets in (57a): *(ng)ajak*. The absence of the AV prefix (i.e. *ajak*) gives rise to a comitative reading only, as shown by reading (i) in (57a); this is a comitative SVC in Balinese. In contrast, the presence of the AV prefix, *ngajak*, leads to an ordinary coordination, which requires a syntactic PIVOT, as in reading (ii) in (57a). The presence of the clausal negator, *tan* 'not', in (57b) forces the coordination structure, which requires SUBJ PIVOT marking. Hence, the presence of an AV prefix on the verb in the second clause is obligatory, as seen in (57b).

(57) Balinese (Shiohara & Arka 2023)

- a. Tiang [mlajah kelompok (ng-)ajak timpal-timpal-e]_{SVC}.
 1SG MID.study group (AV-)invite friend.REDUP-DEF
 (i) ‘I studied in a group together with friends.’ (with *ajak*)
 (ii) ‘I studied in a group and invited friends to join.’ (with *ngajak*)
- b. Tiang [mlajah kelompok], [tan ngajak/*ajak Ketut].
 1SG MID.study group NEG AV.invite/invite Ketut
 ‘I studied in a group, (but) I didn’t invite Ketut (to join).’

In addition, the prosody is different: the SVC in (57a) with the bare verb, *ajak*, has one intonational contour (i.e. without a break), while the coordination in (57b) has a break indicated by a comma after the first VP (cf. prosodic properties of mono-/bi-clausality in Aikhenvald 2006: 7, Dixon 2006: 339, Haspelmath 2016: 308). Likewise, sentence (57a) in its non-SVC or bi-clausal reading (ii) is also accompanied by a prosodic break before the AV verb.

6 Discourse information structure: Contrastive FOCUS and nominalisation

In this final section, we consider the interface between information structure and morphosyntax in AN languages. Recall from Section 3 that contrastive discourse functions are a crucial factor that motivate syntactic variation for fronting. Fronting is of special interest since it involves clefting, which is closely bound with the SUBJ-only restriction on extraction in many AN languages with robust voice systems (cf. Section 4.2.2).

In this section, we look thoroughly at the connection between contrastive DFS and the syntactic structure of clefts from a comparative perspective and demonstrate how cross-linguistic variation can be captured in LFG. We begin by introducing the basic notions of discourse features in information structure (Zaenen 2023 [this volume]) with a primary focus on FOCUS-C because in many AN languages, FOCUS-C is the most common discourse function associated with clefting. Then, we move on to show how FOCUS-C expressions via clefting are structured differently across languages.

The pragmatic uses of clefting in expressing contrastive focus (FOCUS-C) emerge as a motivating factor in the extension of bi-clausal structure across AN languages, as discussed below. In symmetrical-voice languages, bi-clausal clefting is used in combining nominal predicates and headless relative clauses, while

indexing-type languages use mono-clausal clefts without relative constructions. The major difference lies in the gradual erosion of clausal nominalisation. We will see that both types of cleft-structures for expressing FOCUS-C (with/without nominalisation) can be elegantly captured in an LFG analysis to reflect the language-specific variation.

6.1 Information structure: FOCUS-C, fronting and cleft

Topic and focus have long been recognized as discourse functions within information structure. However, this traditional dichotomy view falls short of encompassing all the information structure nuances (Zaenen 2023 [this volume]). Decomposing i-structure features is generally adopted in LFG studies. In our analysis, marked DFS (FOCUS-C/TOPIC-C) are represented by three distinct decomposed features, as demonstrated by Arka & Sedeng (2018) and references therein: contrast, salience, and givenness. The [+contrast] feature is central for FOCUS and is exemplified in (58) below. The [+salient] and [+given] features are typically topic-related, encompassing communicatively important properties, such as the particular frame/entity by which new information should be understood (i.e. the ‘aboutness’ of the topic), and the degree of importance/prominence of one piece of information relative to other bits of information in a given context. The [+salient] feature reflects the speaker’s subjective choice of highlighting one element and making it stand out for communicative purposes. While often closely linked, salience and givenness are distinct: for example, new information, [–given], can be [+salient] (see Riesberg et al. 2018 on information structure across AN languages).

FOCUS-C is a marked FOCUS and is typically characterised by overt marking of the conception of alternatives in the contrastive set it is associated with (cf. Krifka 2008). Clefting is a typical ‘marked’ strategy to express FOCUS-C as seen in the English example of (58a): John is a person in the set of referents associated with the SUBJ (i.e. John, not somebody else). The equivalent structure in Indonesian is given in (58b) below:

- (58) Indonesian (Arka, own knowledge)
- a. It is [John]_{FOCUS-C} [who killed the robber]_{VP:COMMENT|GIVEN}.
 - b. [(Adalah) John]_{PRED/FOCUS-C} [yang membunuh perampok itu]_{SUBJ}.
 be John REL AV-kill robber that
 ‘It’s John who killed the robber.’

Note that English and Indonesian show structural parallelism in their relativisation of the second part of cleft structures, and contrastive FOCUS fronting. Also,

they both show clear evidence of biclausal structures with each part having its own predicate. However, Indonesian *adalah* ‘be’ is optionally present. English requires the empty SUBJ *it*, while Indonesian has no such SUBJ. The fronted NP (*John*) is the predicate, and the (headless) relative with *yang* is actually a (clausal) SUBJ.

6.2 Cross-linguistic variation in fronted FOCUS-C

Fronted content questions in other AN languages of Indonesia also typically employ the same clefting strategy that involves relativisation, including in Indonesian, Sundanese and Sasak as in examples (59)-(61) below. These sentences are biclausal. Note that these languages also allow in-situ mono-clausal content questions with no relativisation required (cf. (59) and (62a) where SUBJ is questioned).

- (59) Indonesian (Arka, own knowledge)
 [Siapa]_{PRED/FOCUS-C} [yang membunuh perampok itu]_{SUBJ}?
 who REL AV-kill robber that
 ‘Who killed the robber?’ (Lit. ‘Who is the one who killed the robber?’)
- (60) Sundanese (Hanafi 1997: 3)
 [Saha?]_{PRED/FOCUS-C} [nu meuli? mobil]_{SUBJ}?
 who REL AV.buy car
 ‘Who bought a car?’
- (61) Menó-Mené Sasak (Arka, fieldwork data)
 [Ape]_{PRED/FOCUS-C} [**(saq)* Amir paling wiq]_{SUBJ}?
 what REL Amir steal yesterday
 ‘What did Amir steal yesterday?’
- (62) Indonesian (Arka, own knowledge)
- [Siapa]_{SUBJ} mem-bunuh perampok itu?
 who AV-kill robber that
 ‘Who killed the robber?’
 - Orang itu membunuh [siapa]_{OBJ}?
 person that AV-kill who
 ‘Who did the person kill?’
 - *Siapa orang itu membunuh ___?

Variation in the above clefting strategies reveals the effect of the SUBJ-only constraint on extraction of FOCUS-C and a change in the constraint in some languages. Philippine/Indonesian type languages with robust grammatical voice cannot front the OBJ Q NP in AV (mono)clauses (Arka 2003a: 27, Kroeger 1993: 50, 208). This is exemplified by the ungrammaticality of (62c) above, in contrast to the acceptable in-situ question (62b). A voice alternation is obligatory in order for OBJ Q NP fronting to be acceptable because it maps the patient onto the SUBJ, and also allows possible clefting of the type seen in (59).

However, in languages where grammatical voice is in decline or has disappeared (as often observed with the erosion of AV verbal morphology), the strict adherence to the SUBJ-only constraint might be eased. This relaxation could allow for the fronting of the OBJ Q NP. However, this can only occur under the condition that the fronted OBJ Q NP necessitates relativization within a bi-clausal structure. Such a phenomenon is evident in Sasak, as demonstrated in (61). Notably, when OBJ Q is fronted, the relativizer *saq* cannot be omitted.

It should be noted that even when the AN verbal voice is completely lost, syntactic voice is not always lost as well. The languages of western and central Flores, such as highly isolating Manggarai and Rongga, exhibit a syntactic passive or undergoer voice without verbal voice morphology (see Arka & Kosmas 2005 for details). The canonical clausal word order in these languages is SVO, and the fronted Q NP also makes use of clefting via relativisation, as seen in Manggarai in (63a) below. In this instance, the fronted Q NP is the actor SUBJ. Despite the absence of AV verbal morphology, the syntactic structure follows the Actor Voice (AV) pattern. Conversely, when the fronted Q NP takes on the role of the undergoer, as depicted in (63b), the structure undergoes an alteration to become Undergoer Voice (UV). Here, the actor is expressed in genitive form, which is characteristic of actor realization in the UV voice within AN languages. Note that the verb form in (63a) is identical to that in (63b). However, they are assigned distinct voice glosses (AV/UV) to signify that they are part of different voice constructions.

(63) Kempo Manggarai (CEMP, Indonesia)

- a. [Cai]_{PRED/FOCUS-C} [ata tengo hau]_{SUBJ}?
 who REL AV.hit you
 ‘Who hit you? (Lit. ‘who is the one hitting you?’) (Semiun 1993: 63)
- b. [Cai]_{PRED/FOCUS-C} [ata tengo gau]?
 who REL UV.hit 2GEN
 ‘Who did you hit?’ (Semiun 1993: 64)

The above discussion has shown how the different morpho-syntactic systems in AN languages are structurally connected in a bi-clausal structure with relativisation. Unlike Philippine/Indonesian type languages, a relaxed constraint on extraction is witnessed in the loss of voice morphology in languages like Mangarai, Flores. In the latter, fronting of a non-SUBJ argument is possible without the need for voice alternation.

The obligatory relativisation in fronted question NPs discussed so far brings us to the important interconnection between FOCUS-C, relativisation, voice and nominalisation. This interconnection is evident in that the AN relative clause used for fronted FOCUS-C is transparently nominal in its structure. Typically, and formally, the relativiser is a nominal phrase marker and thus, the marker is multifunctional. In Tagalog, for instance, the relativiser is the NOM marker for an ordinary NP (see (23)), but also for a verb when its SUBJ is in FOCUS-C in the content question (see (64)). Likewise, marked FOCUS-C in declarative sentences—as seen in Indonesian-type languages like Old Javanese in (65)—also use the same nominalisation strategy through relativisation. The same form *ikang* in (65) is also used as a definite determiner in Old Javanese. The NP flagged by *ang* in Tagalog also receives a definite interpretation. Based on these functional correspondences, we contend that Tagalog *ang*, Old Javanese *ikang*, and the Indonesian pronominal relativiser *yang* are clearly cognates (Kähler 1974: 266–267; Blust 2015: 465; Kaufman 2018: 228–229).

- (64) Tagalog (Kaufman 2018: 219)
 [Sino]_{PRED} [ang d(um)ating]_{SUBJ}?
 who NOM <AV>arrive
 ‘Who arrived?’ (Lit. ‘the coming one is who?’)
- (65) Old Javanese (WMP, Indonesia) (Erawati 2014: 150)
 Ikang naga Taksaka [ikang s-um-ahut wwang atuha-nira].
 DEF dragon Taksaka REL <AV>bite person old-3SG.POSS
 ‘The Taksaka dragon is the one who bit his parent.’

The same pattern of nominalisation involving a fronted FOCUS-C is observed across Philippine-type languages as shown in Table 1. These languages also use the same nominalisation strategy through relativisation. Crucially, there are two morphosyntactic properties worth noting. First, only SUBJ can be fronted as FOCUS-C. Thus, when the transitive patient is in FOCUS-C, the PV must be used, as seen in Table 1. Second, the agent argument of the PV verb is expressed in the genitive, which is the realisation of the possessor in the nominal structure.

Table 1: SUBJ FOCUS-C across the Philippine-type languages (Kaufman 2018: 220)

Tagalog	<i>manga</i>	<i>aŋ</i>	<i>kina:ʔin</i>	<i>naŋ</i>	<i>ba:taʔ</i>
Bikolano	<i>manga</i>	<i>aŋ</i>	<i>kinakan</i>	<i>kan</i>	<i>a:kiʔ</i>
Cebuano	<i>manga</i>	<i>aŋ</i>	<i>ginka:ʔun</i>	<i>han</i>	<i>bataʔ</i>
Hiligaynon	<i>pahuʔ</i>	<i>aŋ</i>	<i>kinaʔun</i>	<i>saŋ</i>	<i>ba:ta</i>
Tausug	<i>mampallam</i>	<i>in</i>	<i>kyaʔun</i>	<i>sin</i>	<i>bataʔ</i>
Ilokano	<i>manga</i>	<i>ti</i>	<i>kinnan</i>	<i>dyay</i>	<i>ubiŋ</i>
Ibanag	<i>manga</i>	<i>ik</i>	<i>kinan</i>	<i>na</i>	<i>abbiŋ</i>
Pangasinan	<i>manga</i>	<i>su</i>	<i>kina</i>	<i>=y</i>	<i>ugaw</i>
Kapampangan	<i>manga</i>	<i>iŋ</i>	<i>pe:ŋa=na</i>	<i>niŋ</i>	<i>anak</i>

[*mango*]_{focus-c} [NOM eat.PV.PFV GEN *child*]_{SUBJ}

‘It was the mango that the child ate.’

(Lit. ‘the mango was the one eaten by the child.’)

The pattern showing the genitive agent in the fronted FOCUS-C with relativisation is also observed in the languages of western Flores, such as Manggarai (Semiun 1993). Recall that Manggarai is highly isolating, but it has a genitive clitic set usable in fronted FOCUS-C questions. Note that the Q *cai* in (66) below is associated with the transitive patient; questioning the agent SUBJ requires no genitive clitic (cf. Kambera example (21) above with (67a) below).

(66) Kempo Manggarai (Semiun 1993)

Cai (ata) tengo gau?

who REL hit 2SG.GEN

‘Who did you hit? (Lit. ‘who is your hitting?’)

In AN languages of the indexing type, the resources for FOCUS-C may also be parasitic to nominalisation/relativisation coding whereby the focused argument ends up being fronted sentence-initially. For example, the Kambera example in (67a) is a content question (FOCUS-C) with equational structure: the verb is affixed with the subject relativiser *ma-* and the verb appears within a nominal (headless) relative clause structure. The nominal article *na* flags the structure as an NP. The same pattern is observed in (67b), where the patient argument is FOCUS-C. Like in Manggarai and Philippine-type languages, the agent in Kambera in (67b) appears as a genitive too, which is the same case as used for the possessor of an NP.

(67) Kambera (Klamer 1998: 132, 318)

- a. Ngga [na ma-palewa-kai] hi mài lai nai?
 who ART RELS-send-2PL.ACC CNJ come LOC DEM
 'Who was the one that send you so that you'd come here?'
- b. [Da kalembi-da]_k [na pa-pa.marihak-na_j nyuna_j]_{SUBJk}.
 ART shirt-3PL.POSS ART RELO-CAUS.be.dirty-3SG he
 'Their shirts_k were (the ones) made dirty by him_j.'

In the languages of Sulawesi, such as Makassarese, where unmarked structures are predicate-initial (like in Philippine-type languages), FOCUS-C formation also requires fronting (Jukes 2006: 341–345). However, while relativisation uses a nominalisation strategy by means of the definite clitic =*a* as seen in (68a), the focus-C formation requires no nominalisation as seen in (68b–c). Makassarese exhibits systematic pronominal indexing, but it still shows the AN voice system. Thus, when agent SUBJ is in FOCUS-C, it requires the homorganic nasal substitution AV prefix on the verb (*aN*- realised as *am*-) as in (68d). Crucially, the sentences with fronted FOCUS-C NPs in (68b–d) are monoclausal.

(68) Makassarese (WMP, Indonesia) (Jukes 2006: 238, 343, 353)

- a. [tau [na=buno=a sorodadu]_{RC}]_{NP}
 person 3=kill=DEF soldier
 'the person killed by a soldier'
- b. Miong=a na=buno kongkong=a.
 cat=DEF 3=kill dog=DEF
 'The dog killed the cat (not something else).'/ 'It's the cat that the dog killed.'
- c. Inai na_i=ba'ji [i Ali]_i?
 who 3=hit PN Ali
 'Who did Ali hit?'
- d. Inai am-ba'ji=i i Udin?
 who AV-hit=3 PN Udin
 'Who hit Udin?'

6.3 Representing information structure in LFG

LFG is well equipped to capture the language-specific variation in fronted FOCUS-C discussed here. There are two kinds of analysis: the (earlier) integrated f-structure analysis (Bresnan & Mchombo 1987, King 1995, among others) and the more recent independent i-structure analysis (Butt & King 1996, Dalrymple & Nikolaeva 2011, among others). In the first analysis, DFs are part of the f-structure and share their values (fully or partially) with argument GFs in the f-structure due to the extended coherence condition (Zaenen 1980) or their anaphoric relation. This analysis is straightforward for cases involving FOCUS-C with no requirement for nominalisation, as in the Makassar examples of (68b–d). Here, the sentences are mono-clausal, and the fronted argument is functionally not the head predicate. For the analysis to work, the sentence-initial XP is identified as FOCUS-C and is licensed by the phrase structure rule shown in (69a) below:

- (69) a. $CP \longrightarrow \begin{array}{c} XP \qquad C' \\ (\uparrow \text{FOCUS-C}) = \downarrow \\ (\uparrow \text{FOCUS-C}) = (\uparrow \text{GF}) \end{array}$
- b. Makassarese Voice Marking:
- i) AV, $aN-$: $(\uparrow \text{FOCUS-C}) = (\uparrow \text{SUBJ})_{\sigma} = (\uparrow_{\sigma} 1:\text{agent})$
 - ii) PV, Clitic_A : $(\uparrow \text{FOCUS-C}) = (\uparrow \text{OBJ})_{\sigma} = (\uparrow_{\sigma} 2:\text{patient})$

The two lines of annotation in (69a) impose a sharing between FOCUS-C and any GF, including adjunct. However, there are also other independent language-specific voice selection constraints given in (69b) to regulate how a core argument is selected as SUBJ/OBJ in Makassarese, particularly when this core argument is also assigned FOCUS-C. Therefore, in light of the rule given in (69b.ii), the example in (68c) (cf. the same example in (70) below) will have the FOCUS-C selected as OBJ. The f-structure is shown in (71) below. We analyse the free NP, which cross-references the agent proclitic, as an adjunct that provides specific information about the agent.

- (70) Makassarese (Jukes 2006: 353)
- inai_j na_i=ba'ji [i Ali]_i?
 who 3=hit PN Ali
 'Who did Ali hit?'

(71) f-structure of sentence (70)

FOCUS-C	[PRED 'PRO'	
	INDEX <i>j</i>	
	PRO-TYPE WH	
Q	[
PRED	'HIT <SUBJ, OBJ>'	
OBJ]	
SUBJ	[PRED 'PRO'	
	PRO-TYPE CL	
	CASE ERG	
	INDEX	[PERS 3]
		NUM SG] <i>i</i>
	{ [PRED 'ALI'	
ADJUNCT	INDEX <i>i</i>	
	N-TYPE PROPER] }	

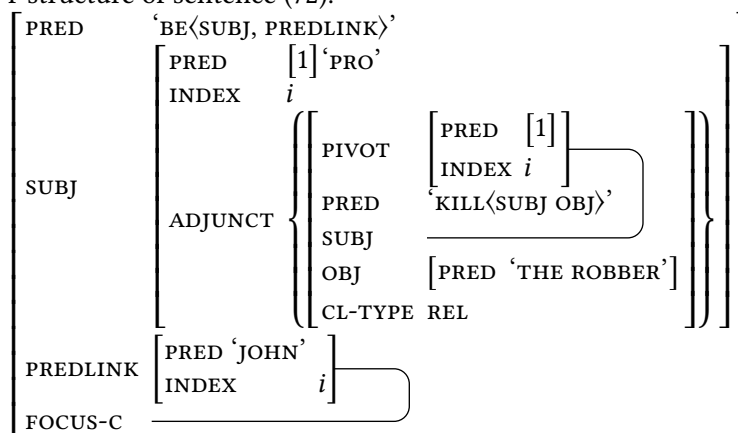
The integrated f-structure analysis just outlined for Makassarese faces an issue when it is applied to fronted FOCUS-C involving bi-clausal or relative clause nominalisation as in Indonesian, as illustrated in (58b), reproduced as (72) below. This is because the FOCUS-C unit is the predicate (cf. the Russian examples discussed by King 1997). One way of resolving this issue is to separate f-structure from i-structure in order to focus on the PRED value only and not its GFs.³³ Since space precludes a full discussion of a separate i-structure analysis in this chapter, we instead demonstrate an integrated f-structure analysis of the fronted Q in Indonesian in (72) below, through the double-tier PREDLINK analysis. This analysis is typically used for the non-verbal predicate with the copula 'be' (Butt et al. 1999, Dalrymple et al. 2004). The (simplified) f-structure in (73) for example (58), repeated in (72), shows that the fronted Q is the FOCUS-C PREDLINK.

(72) Indonesian

[(Adalah) John]_{PRED/FOCUS-C} [yang membunuh perampok itu]_{SUBJ}.
 be John REL AV-kill robber that
 'It's John who killed the robber.'

³³The independent i-structure with a set DF value also allows more than one element in focus. This analysis requires a different DF annotation in the PS rule. The independent i-structure analysis also adopts more sophisticated i-structure conceptions (e.g. with fine-grained distinctions of internal units, such as TOPIC/FOCUS TYPES and BACKGROUND/GIVEN. See King 1997, Dalrymple & Nikolaeva 2011, Butt 2014).

(73) f-structure of sentence (72).



In this analysis, John in (72) is part of the (fronted) PREDLINK (Butt et al. 1999), which is introduced by the copular verb *adalah*.³⁴ In the absence of *adalah*, the analysis specifies the existence of an unpronounced copular verb in the c-structure. The FOCUS-C in the c-structure is occupied by the fronted PREDLINK and so PREDLINK and FOCUS-C share the same value. This connection is signified by the curved lines in (73). Note that the subject is a headless RC marked by *yang*. The headless RC contains [PRED 'PRO'] (tag [1]) supplied by the pronominal relativiser *yang*.³⁵ It is coreferential with the SUBJ/FOCUS-C (i.e. PIVOT) of the RC, and the fronted complement predicate, John (indexed *i*).

The difference between two types of FOCUS-C fronting in indexing-type and symmetrical-voice type languages is captured in LFG by the distinct f-structures in (71) and (73). The f-structure of the indexing type (e.g. Makassarese) in (71) shows a single functional clausal PRED head (i.e. syntactically monoclausal). The Q, *inai* 'who', functions as the question (Q) operator, also identified as FOCUS-C and OBJ (i.e. sharing the same value). In contrast, the f-structure of the Indonesian cleft in (73), which represents the symmetrical-voice type, shows a bi-clausal structure in which the matrix PRED is the copula 'BE' and the embedded relative clause's functional head is 'KILL<SUBJ, OBJ>'. Its SUBJ is identified as FOCUS-C via anaphoric relation (represented by index *i*).

³⁴In LFG, there is more than one way of analysing non-verbal predicates (e.g. nominal predicates) depending on language-specific properties: a single-tier or double-tier analysis. See Andrews (1982), Butt et al. (1999) and Dalrymple et al. (2004) for further discussion.

³⁵[PRED 'PRO'] should be optionally specified in the lexical entry of *yang*. It shows up in the headless RC, but it is not needed in the headed RC as the RC's head noun supplies the PRED value.

To conclude, grammatical variation in FOCUS-C fronting in AN languages can be straightforwardly captured in LFG because of its modular design. In such a design, different dimensions of linguistic information are modelled in separate layers of structure. We have demonstrated how the separation of syntactic representation of linear order (c-structure), relational information about grammatical functions (GFS, or f-structure) and context-related Discourse Function (DF) information (i-structure) makes LFG well suited for explicit linguistic analysis to account for the complex constraints in the interface of morphosyntax and pragmatics.

7 Final remarks

In this chapter, we reviewed a broad range of empirically attested morphosyntactic properties in AN languages. We demonstrated how the parallel correspondence architecture of LFG is used to capture the typological diversity of AN languages at different levels of the grammar. Some of these features have posed descriptive and analytical challenges to traditional grammatical notions. Despite these challenges, LFG emerged as a robust and flexible framework for capturing the dynamics of AN languages' internal grammatical systems and the variation between them. This allows us to account for the AN voice system and related grammatical features in a holistic and coherent way. Further, the application of LFG in AN languages plays a crucial role in increasing the framework's potential to be a well-rounded descriptive and analytical tool for typological and theoretical discussions. Thus, additional documentation of AN languages is expected to uncover richer datasets and linguistic diversity, which will provide an ideal testing ground for LFG's grammar-representing architecture.

Acknowledgments

We wish to thank the anonymous reviewers for their careful and thorough reviews, and we are grateful to Malcolm Ross, Mary Dalrymple and Larry Lovell for their valuable feedback on correcting the information in the earlier draft. We are deeply indebted to Mary Dalrymple for her help on preparing the final version of this chapter in LaTeX.

The first author acknowledges support from the following grants: NSF (BCS-0617198), ARC (DP110100307), and ELDP (IPF0011). The second author acknowledges support from the Australian Government Research Training Program (RTP) and Mr. Wang, the Founder of the Fuguangtang Association, on her research.

Abbreviations

Besides the abbreviations from the Leipzig Glossing Conventions, this chapter uses the following abbreviations.

2P	Second position	MID	Middle Voice
AN	Austronesian	NF	nonfinite
AV	Actor Voice	PMP	Proto-Malayo-
CEMP	Central-Eastern		Polynesian
	Malayo-Polynesian	PN	Proper Name
CN	Common Noun	PREDFOC	Predicate Focus
CNJ	Conjunction	PREP	Preposition
CV	Conveyance Voice	PRT	Particle
DF	Discourse Function	PV	Patient Voice
DOM	Differential Object	RC	Relative Clause
	Marking	REAL	Realis
DV	Dative Voice	REDUP	Reduplication
FOCUS-C	Contrastive Focus	RELO	Object relativizer
GEN	Genitive	RELS	Subject relativizer
GR	Grammatical Relation	SVC	Serial Verb
IN	Inchoative		Construction
IV	Instrumental Voice	UV	Undergoer Voice
LNK	Linker	WMP	Western
LV	Locative Voice		Malayo-Polynesian

References

- Adelaar, Alexander. 1989. Malay influence on Malagasy: Linguistic and culture-historical implications. *Oceanic Linguistics (Special issue on Western Austronesian Languages)* 28(1). 1–46. DOI: 10.2307/3622973.
- Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Serial verb constructions*, 1–62. Oxford: Oxford University Press.
- Aldridge, Edith Catherine. 2004. *Ergativity and word order in Austronesian languages*. Ithaca: Cornell University. (Doctoral dissertation).
- Andersen, Suree & T. David Andersen. 2005. Semantic analysis of the Moronene verbal prefix *moN*. In I Wayan Arka & Malcolm Ross (eds.), *The many faces of Austronesian voice systems: Some new empirical studies*, 243–278. Canberra: Pacific Linguistics. <http://hdl.handle.net/1885/146273>.

- Andréasson, Maia. 2007. The architecture of i-structure. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '07 conference*, 26–43. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/12/papers/lfg07andreasson.pdf>.
- Andrews, Avery D. 1982. The representation of case in Modern Icelandic. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 427–503. Cambridge, MA: The MIT Press.
- Andrews, Avery D. 2023. Clause structure and configurationality. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 375–424. Berlin: Language Science Press. DOI: 10.5281/zenodo.10185952.
- Andrews, Avery D. & Christopher D. Manning. 1999. *Complex predicates and information spreading in LFG*. Stanford: CSLI Publications.
- Arka, I Wayan. 2000. Control and argument structure: Explaining control into subject in Indonesian. Presented at the 4th International Symposium on Malay/Indonesian Linguistics (ISMIL). Jakarta. <http://hdl.handle.net/1885/41056>.
- Arka, I Wayan. 2003a. *Balinese morphosyntax: A lexical-functional approach*. Canberra: Pacific Linguistics. DOI: 10.15144/PL-547.
- Arka, I Wayan. 2003b. Voice systems in the Austronesian languages of Nusantara: Typology, symmetry and Undergoer orientation. *Linguistik Indonesia* 21(1). 113–139.
- Arka, I Wayan. 2006. Voice in Mentawai: A note on its typological position in the Austronesian languages of Indonesia. Presented at the 16th Annual Meeting of the Southeast Asian Linguistics Society at Atma Jaya University, Jakarta, Indonesia.
- Arka, I Wayan. 2011. On modality and finiteness in Indonesian: Complexities of =*nya* nominalisation. In *Proceedings of the International Workshop on TAM and Evidentiality in Indonesian Languages*, 73–89. Tokyo: Tokyo University of Foreign Studies.
- Arka, I Wayan. 2014a. Double and backward control in Indonesian: An LFG analysis. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 26–46. Stanford: CSLI Publications. <http://hdl.handle.net/1885/29065>.
- Arka, I Wayan. 2014b. Locatives and the argument-adjunct distinction in Balinese. *Linguistic Discovery* 12(2). 56–84.
- Arka, I Wayan. 2016. *Bahasa Rongga: Deskripsi, tipologi dan teori [Rongga: Description, typology and theory]*. Jakarta: Penerbit Universitas Katolik Indonesia Atma Jaya.
- Arka, I Wayan. 2017. The core-oblique distinction in some Austronesian languages of Indonesia and beyond. *Linguistik Indonesia* 35(2). 100–142. DOI: 10.26499/li.v35i2.66.

- Arka, I Wayan. 2019. Grammatical relations in Balinese. In Alena Witzlack-Makarevich & Balthasar Bickel (eds.), *Argument selectors: A new perspective on grammatical relations*, 257–299. Amsterdam: John Benjamins. DOI: 10.1075/tsl.123.08ark.
- Arka, I Wayan. 2021. Pivot selection and puzzling relativisation in Indonesian. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 181–202. Oxford: Oxford University Press. DOI: 10.1093/oso/9780192844842.003.0012.
- Arka, I Wayan, Avery D. Andrews, Mary Dalrymple, Meladel Mistica & Jane Simpson. 2009. A linguistic and computational morphosyntactic analysis for the applicative *-i* in Indonesian. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '09 conference*, 85–105. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/14/papers/lfg09arkaetal.pdf>.
- Arka, I Wayan, Mary Dalrymple & Keira Mullan. Forthcoming. Grammatical number in Papuan languages. In Nick Evans & Sebastian Fedden (eds.), *Oxford guide to the Papuan languages*. Oxford: Oxford University Press.
- Arka, I Wayan & Jaladu Kosmas. 2005. Passive without passive morphology? Evidence from Manggarai. In I Wayan Arka & Malcolm Ross (eds.), *The many faces of Austronesian voice systems: Some new empirical studies*, 87–117. Canberra: Pacific Linguistics. <http://hdl.handle.net/1885/146273>.
- Arka, I Wayan & Christopher D. Manning. 2008. Voice and grammatical relations in Indonesian: A new perspective. In Peter K. Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*, 45–69. Stanford: CSLI Publications. <http://hdl.handle.net/1885/26154>.
- Arka, I Wayan & I Nyoman Sedeng. 2018. Information structure in Sembiran Balinese. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in Austronesian languages*, 139–175. Berlin: Language Science Press. DOI: 10.5281/zenodo.1402543.
- Arka, I Wayan & Jane Simpson. 2008. Objective voice and control into subject clauses in Balinese. In Peter Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*, 70–127. Stanford: CSLI Publications. <http://hdl.handle.net/1885/26472>.
- Asikin-Garmager, Eli Scott. 2017. *Sasak voice*. Cedar Rapids: University of Iowa. (Doctoral dissertation). DOI: 10.17077/etd.dfu5gcvv.
- Austin, Peter K. 2004. Clitics in Sasak, eastern Indonesia. Unpublished manuscript, SOAS. <https://www.academia.edu/2993630>.
- Austin, Peter K. & Joan Bresnan. 1996. Non-configurationality in Australian aboriginal languages. *Natural Language & Linguistic Theory* 14(2). 215–268. DOI: 10.1007/bf00133684.

- Bellwood, Peter. 2007. *Prehistory of the Indo-Malaysian archipelago*. Canberra: ANU E-Press. DOI: 10.22459/pima.03.2007. Revised version.
- Belyaev, Oleg. 2023a. Grammatical functions in LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 97–161. Berlin: Language Science Press. DOI: 10.5281/zenodo.10185938.
- Belyaev, Oleg. 2023b. Introduction to LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 3–22. Berlin: Language Science Press. DOI: 10.5281/zenodo.10185934.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics (Volume 1)*, 31–94. Taipei: Academia Sinica.
- Blust, Robert. 2013. *The Austronesian languages*. Revised. Canberra: Asia-Pacific Linguistics. <http://hdl.handle.net/1885/10191>.
- Blust, Robert. 2015. Case-markers of Proto-Austronesian. *Oceanic Linguistics* 54(2). 436–491. DOI: 10.1353/ol.2015.0024.
- Blust, Robert. 2019. The Austronesian homeland and dispersal. *Annual Review of Linguistics* 5(1). 417–434. DOI: 10.1146/annurev-linguistics-011718-012440.
- Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King & John T. III Maxwell. 2010. Second position and the prosody-syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '10 conference*, 106–126. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/15/papers/lfg10boegeletal.pdf>.
- Bowden, John. 2001. *Taba: Description of a south Halmahera language*. Canberra: Pacific Linguistics. DOI: 10.15144/PL-521.
- Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-Functional Syntax*. 2nd edn. (Blackwell Textbooks in Linguistics 16). Malden, MA: Wiley-Blackwell.
- Bresnan, Joan & Jonni M. Kanerva. 1989. Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1). 1–50.
- Bresnan, Joan & Jonni M. Kanerva. 1992. The thematic hierarchy and locative inversion in UG: A reply to Paul Schachter's comments. In Timothy Stowell & Eric Wehrli (eds.), *Syntax and the lexicon*, vol. 26 (Syntax and Semantics), 111–125. San Diego, CA: Academic Press.
- Bresnan, Joan & Ronald M. Kaplan. 1982. Introduction: Grammars as mental representations of language. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, xvii–lii. Cambridge, MA: The MIT Press.

- Bresnan, Joan & Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63(4). 741–782. DOI: 10.2307/415717.
- Brown, Lea. 2001. *A grammar of Nias Selatan*. Sydney: University of Sydney. (Doctoral dissertation). <http://hdl.handle.net/2123/12669>.
- Bugenhagen, Robert D. 1995. *A grammar of Mangap-Mbula: An Austronesian language of Papua New Guinea*. Canberra: Australian National University. (Doctoral dissertation). DOI: 10.15144/PL-C101.
- Butt, Miriam. 2014. Questions and information structure in Urdu/Hindi. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 158–178. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/19/papers/lfg14butt.pdf>.
- Butt, Miriam & Tracy Holloway King. 1991. Semantic case in Urdu. In Lise M. Dobrin, Lynn Nichols & Rosa M. Rodriguez (eds.), *Papers from the 27th regional meeting of the Chicago Linguistic Society*, 31–45.
- Butt, Miriam & Tracy Holloway King. 1996. Structural topic and focus without movement. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '96 conference*. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/LFG1-1996/lfg96butt.pdf>.
- Butt, Miriam & Tracy Holloway King. 2003. Case systems: Beyond structural distinctions. In Ellen Brandner & Heike Zinsmeister (eds.), *New perspectives on case theory*, 53–87. Stanford: CSLI Publications.
- Butt, Miriam, Tracy Holloway King & John T. III Maxwell. 2003. Complex predication via restriction. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '03 conference*, 92–104. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/8/pdfs/lfg03buttetal.pdf>.
- Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A grammar writer's cookbook*. Stanford: CSLI Publications.
- Chang, Anna Hsiou-chuan. 2006. *A reference grammar of Paiwan*. Canberra: Australian National University. (Doctoral dissertation). DOI: 10.25911/5d778712291bf.
- Chang, Yung-Li. 1997. *Voice, case, and agreement in Seediq and Kavalan*. Hsinchu: National Tsing Hua University. (Doctoral dissertation).
- Chang, Yung-Li. 2000. 噶瑪蘭語參考語法 [*A reference grammar of Kavalan*] (Formosan Language Series). Taipei: 遠流 [Yuanliu].
- Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley/Los Angeles: University of California Press.
- Comrie, Bernard. 1978. Ergativity. In Winfred P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 329–394. Austin: University of Texas Press.

- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. 2nd edn. Chicago: University of Chicago Press.
- Croft, William. 2003. Typology. In Mark Aronoff & Janie Rees-Miller (eds.), *The handbook of linguistics*, 337–368. Malden, MA: Blackwell.
- Crowley, Terry. 2002. *Serial verbs in Oceanic: A descriptive typology*. Oxford: Oxford University Press.
- Crowley, Terry. n.d. Grammatical sketch of Enggano. Unpublished manuscript. https://enggano.ling-phil.ox.ac.uk/static/Crowley/Crowley_Grammatical_Sketch_of_Enggano_transcribed.pdf.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar* (Syntax and Semantics 34). New York: Academic Press. DOI: 10.1163/9781849500104.
- Dalrymple, Mary, Helge Dyvik & Tracy Holloway King. 2004. Copular complements: Closed or open? In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '04 conference*, 188–198. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/9/pdfs/lfg04ddk.pdf>.
- Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford reference guide to Lexical Functional Grammar*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198733300.001.0001.
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure* (Cambridge Studies in Linguistics). Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9780511993473.
- Davies, William D. 2010. *A grammar of Madurese*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110224443.
- Davies, William D. & Eri Kurniawan. 2013. Movement and locality in Sundanese wh-questions. *Syntax* 16(2). 111–147. DOI: 10.1111/j.1467-9612.2012.00174.x.
- Davis, Karen. 2003. *A grammar of the Hoava language, Western Solomons*. Canberra: Pacific Linguistics. DOI: 10.15144/PL-535.
- De Guzman, Videia P. 1988. Ergative analysis for Philippine languages: An analysis. In Richard McGinn (ed.), *Studies in Austronesian linguistics*, 323–345. Athens, OH: Ohio University Center for International Studies.
- Dixon, R. M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9781139084987.
- Dixon, R. M. W. 1979. Ergativity. *Language* 55(1). 59–138. DOI: 10.2307/412519.
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge/New York: Cambridge University Press. DOI: 10.1017/cbo9780511611896.
- Dixon, R. M. W. 2006. Serial verb constructions: Conspectus and coda. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Serial verb constructions: A cross-linguistic typology*, 338–350. Oxford: Oxford University Press.

- Donohue, Mark. 2002. Voice in Tukang Besi and the Austronesian focus system. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 81–99. Canberra: Pacific Linguistics. DOI: 10.15144/PL-518.81.
- Donohue, Mark. 2008. Pronouns, clitics, orders, and grammaticalization in Tukang Besi. *Studies in Philippine Languages and Cultures* 17. 1–12.
- Durie, Mark. 1987. Grammatical relations in Acehnese. *Studies in Language* 11(2). 365–399. DOI: 10.1075/sl.11.2.05dur.
- Durie, Mark. 1997. Grammatical structures in verb serialisation. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Complex predicates*. Stanford: CSLI Publications.
- Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1). 1–25.
- Erawati, Ratna. 2014. *Valensi dalam bahasa Jawa Kuna: Suatu kajian morfosintaksis [Valence in Old Javanese: A study of morphosyntax]*. Denpasar: Universitas Udayana. (Doctoral dissertation).
- Erlewine, Michael Yoshitaka. 2018. Extraction and licensing in Toba Batak. *Language* 94(3). 662–697. DOI: 10.1353/lan.2018.0039.
- Falk, Yehuda N. 2000. Pivots and the theory of grammatical functions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '00 conference*, 122–138. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/LFG5-2000/pdfs/lfg00falk.pdf>.
- Falk, Yehuda N. 2006. *Subjects and universal grammar: An explanatory theory*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9780511486265.
- Falk, Yehuda N. 2007. Constituent structure and GFs in the Hebrew nominal phrase. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Chris Manning (eds.), *Architectures, rules, and preferences: Variations on themes by Joan W. Bresnan*, 103–126. Stanford: CSLI Publications.
- Findlay, Jamie Y., Roxanne Taylor & Anna Kibort. 2023. Argument structure and mapping theory. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 699–778. Berlin: Language Science Press. DOI: 10.5281/zenodo.10185966.
- Foley, William A. 1998. Symmetrical voice systems and precategoriality in Philippine languages. Paper presented at the LFG '98 Conference, University of Queensland. Brisbane, Australia.
- Foley, William A. & Robert Van Valin Jr. 1984. *Functional syntax and universal grammar*. Cambridge, UK: Cambridge University Press.
- François, Alexandre. 2006. Serial verb constructions in Mwotlap. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Serial verb constructions*, 223–238. Oxford: Oxford University Press.

- Georgopoulos, Carol. 1991. *Syntactic variables: Resumptive pronouns and A' binding in Palauan*. Dordrecht: Kluwer Academic Publishers.
- Givón, Talmy. 2001. *Syntax: An introduction*, vol. I. Amsterdam & Philadelphia: John Benjamins.
- Halpern, Aaron. 1995. *On the placement and morphology of clitics*. Stanford: CSLI Publications.
- Halpern, Aaron & Arnold M. Zwicky (eds.). 1996. *Approaching second: Second position clitics and related phenomena*. Stanford: CSLI Publications.
- Hanafi, Nurachman. 1997. *A typological study of Sundanese*. Melbourne: La Trobe University. (Doctoral dissertation).
- Haspelmath, Martin. 2007. Ditransitive alignment splits and inverse alignment. *Functions of Language* 14(1). 79–102. DOI: 10.1075/foL.14.1.06has.
- Haspelmath, Martin. 2016. The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics* 17(3). 291–319. DOI: 10.1177/2397002215626895.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780199252695.001.0001.
- Heath, Jeffrey. 1975. Some functional relationships in grammar. *Language* 51(1). 89–104. DOI: 10.2307/413151.
- Hemmings, Charlotte. 2021. Subjects in Austronesian: Evidence from Kelabit. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 160–180. Oxford: Oxford University Press. DOI: 10.1093/oso/9780192844842.003.0011.
- Hemmings, Charlotte. In preparation. Verbs in contemporary Enggano. In Mary Dalrymple, I Wayan Arka & Charlotte Hemmings (eds.), *Enggano: Historical and contemporary perspectives*. <https://enggano.ling-phil.ox.ac.uk/static/papers.html>.
- Hill, Deborah. 2002. Longgu. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic languages*, 538–561. Richmond, Surrey: Curzon Press.
- Himmelmänn, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. New York: Routledge.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299. DOI: 10.2307/413757.
- Hsin, Aili. 1996. Noun phrase structure and focus marking in Kavalan. *Tsing Hua Journal of Chinese Studies (New Series)* 26(3). 323–364.

- Hyslop, Catriona. 2001. *The Lolovoli dialect of the North-East Ambae language, Vanuatu*. Canberra: Australian National University. (Doctoral dissertation). DOI: 10.15144/PL-515.
- Indrawati, Ni Luh Ketut Mas. 2014. The typological perspective of the Balinese serial verb constructions. In I Wayan Arka & Ni Luh Ketut Mas Indrawati (eds.), *Argument realisations and related constructions in Austronesian languages: Papers from 12-ICAL*, vol. 2, 353–368. Canberra: Asia-Pacific Linguistics. <http://hdl.handle.net/1885/12059>.
- Jespersen, Otto. 1924. *The philosophy of grammar*. London: Allen & Unwin. DOI: 10.4324/9780203716045.
- Jukes, Anthony. 2006. *Makassarese (basa Mangkasara’): A description of an Austronesian language of South Sulawesi*. Melbourne: University of Melbourne. (Doctoral dissertation). <https://www.oxis.org/theses/jukes-2006.pdf>.
- Kähler, Hans. 1974. Relative clause formation in some Austronesian languages. *Oceanic Linguistics* 13(1/2). 257–277. DOI: 10.2307/3622746.
- Kaplan, Ronald M. & Jürgen Wedekind. 1993. Restriction and correspondence-based translation. In *Proceedings of the 6th conference of the European chapter of the ACL (EACL 1993)*, 193–202. Association for Computational Linguistics. DOI: 10.3115/976744.976768.
- Katagiri, Masumi. 2005. Voice, ergativity and transitivity in Tagalog and other Philippine languages: A typological perspective. In I Wayan Arka & Malcolm Ross (eds.), *The many faces of Austronesian voice systems: Some new empirical studies*. Canberra: Pacific Linguistics. <http://hdl.handle.net/1885/146273>.
- Kaufman, Daniel. 2018. Austronesian predication and the emergence of biclausal clefts in Indonesian languages. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in Austronesian languages*, 207–246. Berlin: Language Science Press. DOI: 10.5281/zenodo.1402547.
- King, Tracy Holloway. 1995. *Configuring topic and focus in Russian*. Stanford: CSLI Publications.
- King, Tracy Holloway. 1997. Focus domains and information structure. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG ’97 conference*, 2–13. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/LFG2-1997/lfg97king.pdf>.
- King, Tracy Holloway. 2005. Cliticizing LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG ’05 conference*. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/10/pdfs/lfg05king.pdf>.
- Klamer, Marian. 1996. Kambera has no passive. In Marian Klamer (ed.), *Voice in Austronesian*, vol. 39 (NUSA-linguistic Studies of Indonesian and Other Languages in Indonesia), 12–30. Jakarta: Universitas Atma Jaya.

- Klamer, Marian. 1998. *A grammar of Kambera* (Mouton Grammar Library). Berlin: Mouton de Gruyter. DOI: 10.1515/9783110805536.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4). 243–276. DOI: 10.1556/aling.55.2008.3-4.2.
- Kroeger, Paul R. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Kroeger, Paul R. 2023. The non-Malayic languages of Borneo. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia*. Oxford: Oxford University Press.
- Lenggang, Zainuddin H. R., Be Kim Hoa Nio, Mohd. Ansyar Zainil & Syofyan Adam. 1978. *Bahasa mentawai*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Li, Paul Jen-kuei & Shigeru Tsuchida. 2001. *Pazih dictionary* (Language and Linguistics Monograph Series A2). Taipei: Academia Sinica.
- Li, Paul Jen-kuei & Shigeru Tsuchida. 2002. *Pazih texts and songs*. Taipei: Academia Sinica.
- Liu, (Adlay) Kun-Long. 2017. *Syntactic interactions with information structure in Squliq Atayal*. Canberra: Australian National University. (Doctoral dissertation). DOI: 10.25911/5d7632016f800.
- Lynch, John, Malcolm Ross & Terry Crowley (eds.). 2002. *The Oceanic languages*. Richmond, Surrey: Curzon Press.
- Manning, Christopher D. 1994. *Ergativity: Argument structure and grammatical relations*. Stanford: Stanford University. (Doctoral dissertation). <https://nlp.stanford.edu/~manning/dissertations/Manning-Christopher-ergativity-1994.pdf>.
- Natarina, Ari. 2018. *Complementation in Balinese: Typological, syntactic and cognitive perspectives*. Cedar Rapids: University of Iowa. (Doctoral dissertation). DOI: 10.17077/etd.1h90npxe.
- Norwood, Clodagh Margaret. 2002. *Voice and valency in Karo Batak*. Melbourne: La Trobe University. (Doctoral dissertation).
- O'Connor, Robert. 2002. Clitics and phrasal affixation in constructive morphology. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '02 conference*, 315–332. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/7/pdfs/lfg02oconnor.pdf>.
- Palmer, Bill. 1999. *A grammar of the Kokota language, Santa Isabel, Solomon Islands*. Sydney: University of Sydney. (Doctoral dissertation). <https://www.smg.surrey.ac.uk/media/1128/prel.pdf>.

- Pawley, Andrew. 2002. The Austronesian dispersal: Languages, technologies, people. In Peter Bellwood & Colin Renfrew (eds.), *Examining the farming/language dispersal hypothesis*, 251–273. Cambridge, UK: McDonald Institute for Archaeological Research.
- Pawley, Andrew & Medina Pawley. 1998. Canoes and seafaring. In Malcolm Ross, Andrew Pawley & Meredith Osmond (eds.), *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society (Volume 1: Material Culture)*, 173–209. Canberra: Pacific Linguistics. DOI: 10.15144/PL-C152.173.
- Rákosi, György. 2023. Anaphora. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 165–192. Berlin: Language Science Press. DOI: 10.5281/zenodo.10185940.
- Riesberg, Sonja. 2014. *Symmetrical voice and linking in Western Austronesian*. Berlin: Mouton de Gruyter.
- Riesberg, Sonja, Kurt Malcher & Nikolaus P. Himmelmann. 2019. How universal is agent-first? Evidence from symmetrical voice languages. *Language* 95(3). 523–561. DOI: 10.1353/lan.2019.0055.
- Riesberg, Sonja, Asako Shiohara & Atsuko Utsumi (eds.). 2018. *Perspectives on information structure in Austronesian languages* (Studies in Diversity Linguistics). Berlin: Language Science Press. DOI: 10.5281/zenodo.1402571.
- Sadler, Louisa. 2023. LFG and Semitic languages. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 1651–1702. Berlin: Language Science Press. DOI: 10.5281/zenodo.10186020.
- Sag, Ivan A. & Carl Pollard. 1991. An integrated theory of complement control. *Language* 67. 63–113. DOI: 10.2307/415539.
- Sawaki, Yusuf. 2016. *A grammar of Woi: An Austronesian language of Yapen Island, Western New Guinea*. Canberra: Australian National University. (Doctoral dissertation). DOI: 10.25911/5d6c3fd1c9f50.
- Schachter, Paul. 1977. Reference-related and role-related properties of subjects. In Peter Cole & Jerrold M. Sadock (eds.), *Grammatical relations* (Syntax and Semantics 8), 279–306. New York: Academic Press.
- Sells, Peter. 2023. LFG and Minimalism. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 1919–1965. Berlin: Language Science Press. DOI: 10.5281/zenodo.10186044.
- Semiun, Agustinus. 1993. *The basic grammar of Manggarai: Kempo subdialect*. Melbourne: La Trobe University. (MA thesis).
- Shen, Chia-chi. 2005. *Reflexives and reciprocals in Kavalan*. Taipei: National Taiwan University. (MA thesis). DOI: 10.6342/NTU.2005.00362.

- Shiohara, Asako & I Wayan Arka. 2023. Balinese, Sasak, and Sumbawa. In Alexander Adelaar & Antoinette Schapper (eds.), *Oxford guide to the Malayo-Polynesian languages of south east Asia*. Oxford: Oxford University Press.
- Skoglund, Pontus, Cosimo Posth, Kendra Sirak, Matthew Spriggs, Frederique Valentin, Stuart Bedford, Geoffrey R. Clark, et al. 2016. Genomic insights into the peopling of the Southwest Pacific. *Nature* 538(7626). 510–513. DOI: 10.1038/nature19844.
- Taylor, John. 2003. *Linguistic categorization*. 3rd edn. Oxford: Oxford University Press.
- Teng, Stacy Fang-Ching. 2005. Grammatical relations in Puyuma. In I Wayan Arka & Malcolm Ross (eds.), *The many faces of Austronesian voice systems: Some new empirical studies*, 1st edn., 137–152. Canberra: Pacific Linguistics. <http://ir.sinica.edu.tw/handle/201000000A/57193>.
- Teng, Stacy Fang-Ching. 2008. *A reference grammar of Puyuma*. Canberra: Pacific Linguistics. <http://ir.sinica.edu.tw/handle/201000000A/57143>.
- Utsumi, Atsuko. 2013. Talaud verbs: Paradigm of basic verbs. *Tokyo University Linguistic Papers* 33. 319–361.
- Van Valin, Robert, Jr. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning, and function*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9781139166799.
- Vincent, Nigel. 2023. Raising and control. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 603–647. Berlin: Language Science Press. DOI: 10.5281/zenodo.10185962.
- von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19. 245–274. DOI: 10.1093/jos/19.3.245.
- Wei, Ting-Chi. 2009. 阿美語疑問詞研究 [Wh-interrogatives in Amis]. *Language and Linguistics* 10(2). 315–374.
- Wouk, Fay. 2002. Voice in the languages of Nusa Tenggara Barat. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 285–309. Canberra: Pacific Linguistics. DOI: 10.15144/PL-518.285.
- Wu, Jing-lan Joy. 2006. *Verb classification, case marking, and grammatical relations in Amis*. State University of New York at Buffalo. (Doctoral dissertation).
- Yeh, Li-Chen. 2019. The mismatch between morphological symmetry and syntactic ergativity in Paze. In Miriam Butt, Tracy Holloway King & Ida Toivonen (eds.), *Proceedings of the LFG '19 conference*, 415–430. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/LFG-2019/lfg2019-yeh.pdf>.

- Yeh, Li-Chen. In preparation. *Social-historical dynamics of contact and grammatical changes in Pazeh-Kaxabu*. Australian National University. (Doctoral dissertation).
- Zaenen, Annie. 1980. *Extraction rules in Icelandic*. Reprinted by Garland Press, New York, 1985. Harvard University. (Doctoral dissertation).
- Zaenen, Annie. 1993. Unaccusativity in Dutch: Integrating syntax and lexical semantics. In James Pustejovsky (ed.), *Semantics and the lexicon*, 129–161. Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-011-1972-6_9.
- Zaenen, Annie. 2023. Information structure. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 823–853. Berlin: Language Science Press. DOI: 10.5281/zenodo.10185972.

