Handbook of Lexical Functional Grammar

Edited by

Mary Dalrymple

Empirically Oriented Theoretical
Morphology and Syntax

Empirically Oriented Theoretical Morphology and Syntax

Chief Editor: Stefan Müller
Consulting Editors: Berthold Crysmann, Ana R. Luís

In this series:

1. Lichte, Timm. Syntax und Valenz: Zur Modellierung kohärenter und elliptischer Strukturen mit Baumadjunktionsgrammatiken.

2. Bîlbîie, Gabriela. Grammaire des constructions elliptiques: Une étude comparative des phrases sans verbe en roumain et en français.

3. Bowern, Claire, Laurence Horn & Raffaella Zanuttini (eds.). On looking into words (and beyond): Structures, Relations, Analyses.

4. Bonami, Olivier, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer. The lexeme in descriptive and theoretical morphology.

5. Guzmán Naranjo, Matías. Analogical classification in formal grammar.

6. Flick, Johanna. Die Entwicklung des Definitartikels im Althochdeutschen: Eine kognitiv-linguistische Korpusuntersuchung.

7. Crysmann, Berthold & Manfred Sailer (eds.). One-to-many relations in morphology, syntax, and semantics.

8. Zinova, Yulia. Russian verbal prefixation: A frame semantic analysis.

9. Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig. Head-Driven Phrase Structure Grammar: The handbook.

10. Diewald, Gabriele & Katja Politt (eds.). Paradigms regained: Theoretical and empirical arguments for the reassessment of the notion of paradigm.

11. Freywald, Ulrike, Horst J. Simon & Stefan Müller (eds.). Headedness and/or grammatical anarchy?

Handbook of Lexical Functional Grammar

Edited by

Mary Dalrymple

Mary Dalrymple (ed.). 2023. *Handbook of Lexical Functional Grammar*
(Empirically Oriented Theoretical Morphology and Syntax). Berlin: Language
Science Press.

Contents

Part I

Overview and introduction

Chapter 1

Grammatical functions in LFG

Oleg Belyaev

Lomonosov Moscow State University, Institute of Linguistics of the Russian Academy of Sciences, and Pushkin State Russian Language Institute

Grammatical functions (GFs) such as subject and object play a central role in the architecture of LFG, which makes it quite different from most other formal theories of grammar. In this chapter, I discuss the motivation behind this design decision and the ways in which grammatical functions are distinct from each other: their classification and the properties of certain individual GFs, namely subjects, sentential complements (COMP) and possessors. I also discuss the status of so-called overlay or discourse functions, which serve to specify the status of GFs with respect to additional syntactic constraints.

1 Introduction

LFG is different from most other frameworks in that grammatical functions (GFs) are first-class citizens of syntactic structure. The set of available GFs is viewed as universal, and each GF is associated with a distinct set of structural properties. Some syntactic rules and generalizations refer to individual GFs directly; others refer to their relative ranking, but, unlike GB/Minimalism (Sells forthcoming [this volume]) or HPSG (Przepiórkowski forthcoming [this volume]), the ranking itself is directly stipulated and is secondary to grammatical function status. The list of grammatical functions used in most LFG work includes subjects (SUBJ), direct objects (OBJ), secondary objects (OBJ$_\theta$ or OBJ2), obliques (OBL$_\theta$), and adjuncts (ADJ), which are familiar from traditional grammar but given more exact definitions in LFG. This list is not arbitrary; it is motivated by the classification of grammatical functions into ungovernable (ADJ) vs. governable functions, terms (SUBJ and OBJ) vs. non-terms, semantically restricted (OBJ$_\theta$ and OBL$_\theta$) vs. unrestricted; each class is associated with a distinct expected pattern of behaviour.

Oleg Belyaev

The list of basic GFs is also motivated by the regularity of mapping between semantic roles and their syntactic expression: the cross-classification of GFs into two binary features [±o] and [±r] and the mapping principles assumed in Lexical Mapping Theory (Bresnan & Kanerva 1989) correctly predict both the regular mappings and their possible permutations. More unusually, LFG also treats certain specialized grammatical functions – namely, clausal complements (COMP), possessors (POSS) and sometimes nonverbal predicates (PREDLINK) – as theoretical primitives on par with subjects and objects.

LFG also uses so-called overlay functions to represent the locus of long-distance dependencies like wh-extraction. These do not formally belong to the class of grammatical functions, but are similar in that they are occupied by the same f-structures that represent clausal participants. In earlier versions of LFG, most overlay functions were called "discourse functions" and also represented information structure notions such as topic and focus. In modern LFG, there is usually a separate level for information structure, and there is no need to duplicate it at f-structure. Instead, a single function, here called DIS, is used for all long-distance dependencies; some authors postulate additional overlay functions to model other grammatical information, such as PIVOT for "pivots" in Falk (2006). To the extent that overlay functions are related to grammatical functions, they will be discussed in this chapter; further information on overlay functions with respect to long-distance dependencies is found in Kaplan & Zaenen forthcoming [this volume].

In this chapter, I summarize the key elements of the LFG understanding of grammatical functions. In Section 2, I briefly discuss the formal status of grammatical functions and their role as syntactic primitives in LFG. In Section 3, I describe the main mechanism through which grammatical functions obtain their distinctive properties – their hierarchical ordering and cross-classification. In Section 4, I turn to individual grammatical functions – subjects, objects and obliques – and discuss their distinctive properties that do not follow from their classification or ranking in the hierarchy. Finally, in Section 5, I discuss so-called overlay functions, which represent additional functions that link clausal participants to the wider sentential or discourse context.

2 General concepts

Grammatical functions in LFG represent all kinds of relations that syntactic dependents may have to their predicates. This includes both grammatical relations like subject, object or adjunct and additional functions – so-called overlay functions – that situate the event participant in some wider cross-clausal or discourse

context (e.g. DIS for dislocated – usually topicalized or focalized – elements, or RELPRO for relative pronouns). The values of grammatical functions are normally also event participants – thus, in the words of Bresnan et al. (2016), grammatical functions can be called "the "relators" of c-structure to a[rgument]-structure" (p. 94). However, this is not always the case: adjuncts do not appear at argument-structure, and expletive arguments, like *it* in *It rains*, are purely syntactic and do not correspond to any semantic argument.

In formal terms, a GRAMMATICAL FUNCTION is any f-structure attribute that has an f-structure as its value[1] and whose occurence is governed by Completeness, Coherence and Extended Coherence. Completeness requires that features listed as arguments in a PRED feature value appear within the same f-structure as this PRED. Coherence prevents governable GFs (see Section 3.3) from appearing in f-structures where they are not listed in the PRED value. Extended Coherence restricts the occurence of *ungovernable* GFs: adjuncts and overlay functions. Adjuncts can only appear in f-structures that have a PRED feature (regardless of its value), while overlay functions like DIS (for dislocated constituents), RELPRO (relative pronouns), TOPIC and FOCUS (see Section 5) must be linked to non-overlay functions through structure sharing or anaphora.

For example, (1) represents the f-structure of the sentence 'Peter met Paul in Rome'. The value of the feature PRED includes, in angled brackets, the list of arguments that are required by the verb 'meet' – in English, this is a transitive verb that selects a subject and an object. These arguments appear as the features SUBJ and OBJ that have f-structures representing the NPs *Peter* and *Paul* as their values. The PP *in Rome* is not selected by the verb (its occurrence is not obligatory) and is represented as an element of the set-valued feature ADJ, for adjunct. The preposition *in*, which contributes semantic content, has its own f-structure with the feature PRED whose value defines a valency for OBJ. The nouns *Peter*, *Paul* and *Rome* do not require any syntactic arguments, and hence their PRED feature values lack a list of arguments in angle brackets. For more detail on how arguments and adjuncts are licensed at f-structure, see Belyaev forthcoming(a) [this volume].

---

[1]F-structures appear as values not exclusively with GFs. For example, many authors, among others Alsina & Vigo (2014) and Haug & Nikitina (2015), use the function AGR as a "bundle" of agreement features that is an f-structure that never has a PRED value and that is neither an argument nor an adjunct.

Oleg Belyaev

$$(1) \quad \begin{bmatrix} \text{PRED} & \text{'MEET} \langle \text{SUBJ, OBJ} \rangle \text{'} \\ \text{TENSE} & \text{PAST} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PETER'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'PAUL'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'IN} \langle \text{OBJ} \rangle \text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'ROME'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

The fact that such dependents are represented as values of f-structure *features* is not at all trivial. The term "grammatical relations" used in typology implies that arguments and clauses are viewed as *objects* literally linked to each other via *relations*. Thus, where LFG has (f SUBJ) = g (f is a function, SUBJ is an argument, g is the feature value), the intuitive tradition would rather have SUBJ(f) = g (SUBJ is a function, f is an argument, g is the value). The LFG view has certain interesting consequences for the handling of many syntactic phenomena. For example, the "Coordinate Structure Constraint" (Ross 1967) has no special status in the framework – its effects are of exactly the same nature as the scoping of grammatical features (such as mood or case) over conjuncts in coordinate structures. This is a direct consequence of the fact that grammatical functions such as SUBJ or OBJ are features in exactly the same sense as grammatical features such as CASE or MOOD; for more information, see Patejuk forthcoming [this volume].

A core tenet of LFG is that grammatical functions are theoretical primitives; their set is universal and their properties are not derived from other, more fundamental principles.[2]

The LFG view of GFs as primitives amounts to saying that neither phrase structure relations nor semantics are sufficient to account for all the properties of individual arguments. As discussed in Belyaev forthcoming(c) [this volume] and Andrews forthcoming [this volume], the mapping from c-structure to grammatical functions is relatively unconstrained. X′ Theory, in formulations like that

---

[2]Lexical Mapping Theory (Bresnan & Kanerva 1989) is sometimes interpreted as involving the decomposition of grammatical functions into bundles of two binary features: [±r], [±o], cf. e.g.: "Basic argument functions are not atomic but decomposable into features" (Kibort 2014). Under this view, it is these features that are primitives, instead of GFs. But lexical mapping theory can also be interpreted as a classification rather than an actual decomposition; this is the position taken, for example, in the Oxford Reference Guide to LFG (Dalrymple et al. 2019).

of Bresnan (2001) and Bresnan et al. (2016), does impose certain restrictions, but these are very general and do not impose any specific mapping. For example, it is assumed that complements of lexical projections map to grammatical functions, but no specific mapping is enforced: the complement of VP does not have to map to OBJ, but can map to any grammatical function, even SUBJ. Thus in King (1995), all postverbal (contrastive) foci in Russian, including subjects (2), are analyzed as VP complements.

(2) Russian (Slavic > Indo-European)
Evgenija Onegina napisal **Puškin**
E.:ACC   O.:ACC   wrote   P.:NOM
'It was **Pushkin** who wrote "Eugene Onegin".'

IP
NP
(↑ OBJ) = ↓
(↑ TOPIC) = ↓
Evgenija Onegina
I′
↑=↓
I
↑=↓
napisal
VP
↑=↓
NP
(↑ SUBJ)=↓
Puškin
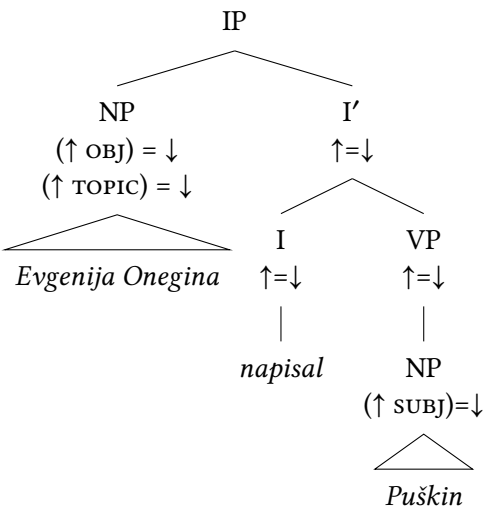
In fact, a consistent mapping cannot be assumed even in so-called configurational languages like English: while in English declarative sentences, objects appear in Comp of VP, the arrangement changes in interrogative sentences, where objects occupy the clause-initial position (Spec of CP or CP adjunct) but the Comp of VP is left empty. Since LFG uses no transformations or any similar mechanism, this has to be accounted for by positing a notion of grammatical function independent from c-structure position.

Grammatical functions are also distinct from semantic roles. A patient, for example, may map to either OBJ (in the active voice) or SUBJ (in the passive), as evidenced by its syntactic properties (e.g., control of verb agreement, reflexive binding). In LFG, these two sentence types are defined as two different *lexical mappings* between semantic roles and GFs. While in terms of argument structure,

i.e. the mapping from semantic roles to GFs, the passive is treated as derivative to the active, at f-structure passive subjects are genuine, first-class subjects that are not derived from objects in any sense.

Finally, grammatical functions cannot be equated to case marking or another argument encoding mechanism, such as verb agreement. First of all, there are many languages which completely lack both agreement and case marking, but which nevertheless display evidence for grammatical functions. Thus Mandinka (Mande > Niger-Congo), which lacks both case marking and verbal indexing, nevertheless displays a distinction between the subject (sole argument of intransitive verbs, i.e. S in typological terminology, and the agent of transitive verbs, i.e. A) and all other arguments in a number of different constructions (Creissels 2019). For instance, pronominal resumption in relative clauses is only available for non-subject arguments. In (3a) and (3b), subjects (S and A arguments, respectively) are relativized, and the resumptive pronoun *à* cannot appear in the subordinate clause in the normal subject position; the subject is represented by a gap. In contrast, in (3c), it is the object that is relativized, and the pronoun *à* may (optionally) appear in the object position after the verb.

(3) Mandinka (Mande > Niger-Congo: Creissels 2019: 339)

    a. S relativized: resumption ungrammatical
        mùs-ôo   **míŋ** (*à) táa-tá   fàr-ôo   tó
        woman-DET REL   3SG go-COMPL.POS rice.field-DET LOC
        ('the woman **who** went to the rice field')

    b. A relativized: resumption ungrammatical
        mùs-ôo   **míŋ** (*à) yè   fǎaŋ-ó   tǎa
        woman-DET REL   3SG COMPL.POS cutlass-DET take
        ('the woman who took the cutlass')

    c. P relativized: resumption possible
        fǎaŋ-ò   **míŋ** mùs-ôo   yè   à   tǎa
        cutlass-DET REL woman-DET COMPL.POS 3SG take
        'the cutlass that the woman took'

Furthermore, case marking or agreement do not always consistently identify specific grammatical functions. For example, in Icelandic (Andrews 1982) agreement is always with the nominative argument, but subjects can be non-nominative. Many languages with differential object marking (DOM) allow nominative objects (Dalrymple & Nikolaeva 2011). For example, in Ossetic human objects are normally genitive-marked (4a) and inanimate objects are nominative-marked (4b), i.e. the case marking of subjects and objects can be identical.

(4) Ossetic (Iranian > Indo-European)

 a. Human P: genitive
     alan **šošlan-ə** fetː-a
         S.-GEN   see.PFV-PST.3SG
     'Alan saw **Soslan**.'

 b. Inanimate P: nominative
     alan **šťol** fetː-a
         table see.PFV-PST.3SG
     'Alan saw **a/the table**.'

Of course, this is not to say that grammatical functions never systematically correspond to any syntactic or morphological marking; if they didn't, there would be no means of identifying them. The point of treating grammatical functions as primitives is that we *cannot*, as a general rule, reduce them to any other linguistic phenomena such as case marking or word order. This logic is in line with the general spirit of LFG, which can be termed "anti-reductionist" in that it strives to factorize grammatical phenomena into distinct notions responsible for distinct patterns of behaviour, which may or may not correlate systematically across languages. Thus, in the LFG treatment of argument encoding, constituent structure, semantic roles, and case marking are all formally independent from each other. The framework itself puts no constraints on their relationship; it is the task of the theorist is to establish how exactly they can or cannot correlate, both cross-linguistically and within individual languages.

We also have to assume, as a working hypothesis, that individual grammatical functions are associated with core sets of syntactic properties that are relatively stable across languages. If this is not the case, then using such terms as "subject" or "direct object" as anything more than convenient language-internal labels is not justified. This issue is still at the centre of much typological discussion, cf. the overview in Bickel (2010). Theoretical discussion is mostly limited to LFG as the only major framework where grammatical relations play a central role. In other frameworks, grammatical functions only exist, at best, in the form of an ordering relation among arguments – this is true at least for most variants of HPSG and Simpler Syntax (Culicover & Jackendoff 2005, Varaschin forthcoming [this volume]). Thus, in recent versions of HPSG there is a list ARG-ST (or DEPS) containing all verbal arguments; the subject is the first element of this list, the direct object, the second, and so on, generally according to the Keenan–Comrie hierarchy (Keenan & Comrie 1977). In many instances, both approaches make the same predictions, because in LFG the GF hierarchy also plays a major role (see Section 3.2);

Oleg Belyaev

for example, in both LFG (Rákosi forthcoming [this volume]) and HPSG (Müller & Branco 2021), anaphoric relations are licensed by the relative ranking of verbal arguments. But the key practical difference is that in HPSG or Simpler Syntax, distinctive properties are not associated with individual grammatical functions. For example, in LFG it is possible to analyze sentences as having only a subject (subj) and a secondary object (obj$_\theta$, without a primary obj) when the "second-ranking" argument is deemed to lack features commonly associated with direct objects. This is done, for example, for certain classes of predicates in Plains Cree (Dahlstrom 2009) and for unmarked direct objects in differential object marking systems in the analysis of Dalrymple & Nikolaeva (2011). Even subjectless sentences are possible if the highest-ranking argument lacks properties that are associated with subjecthood (Kibort 2006). The standard LFG analysis of complementation (Section 4.3) also relies on the grammatical functions comp and xcomp (for clausal complements) being distinct from obj. All of this would be impossible if grammatical functions were just an issue of ranking.

While gfs have been a cornerstone of LFG since its inception, a variant of this framework without the traditional notion of gf is also conceivable. Such an attempt was made in Patejuk & Przepiórkowski (2016), who propose replacing features such as subj, obj and adjunct with an ordered set deps in the style of HPSG. A detailed counterargument to this proposal can be found in Kaplan (2017).

In the following sections, I will describe the standard view of grammatical functions in current LFG: their inventory, their classification, and the properties of the core grammatical functions.

3 The classification of grammatical functions

3.1 General remarks

LFG generally operates with the following set of grammatical functions (with the addition of overlay functions, which will be discussed in Section 5):

(5) subj   subject
     obj   object
     obj$_\theta$ secondary object
     obl$_\theta$ oblique
     comp (xcomp)   complement (closed/open)
     predlink   nonverbal predicate in copular constructions
     adj (xadj)   adjunct (closed/open)
     poss   possessor

10

The θ in OBJ$_\theta$ and OBL$_\theta$ represents the particular semantic role that is filled by the argument. For example, a secondary object and an oblique with the semantic role Goal will be called OBJ$_{GOAL}$ and OBL$_{GOAL}$, respectively. Thus OBJ$_\theta$ and OBL$_\theta$ are not individual GFs but "families" of GFs associated with particular semantic roles, but sharing some common properties. The main motivation for this will be discussed in Section 3.5.

As discussed above, GFs in LFG are theoretical primitives on a par with such entities as constituents, morphosyntactic or phonological features. Such primitives are never given definitions or identified on the basis of a fixed set of tests or criteria; rather, they are associated with a set of properties and used as building blocks for hypotheses whose predictions are to be tested. But this does not mean that the list of GFs in (5) is completely arbitrary. On the contrary, in the following sections I will show how the core GFs (SUBJ, OBJ, OBJ$_\theta$, OBL$_\theta$, ADJ) are mostly distinguished on the basis of three classifications: ungovernable (ADJ) vs. governable, term (SUBJ, OBJ, OBJ$_\theta$) vs. non-term, semantically unrestricted (SUBJ, OBJ) vs. restricted. This only leaves the distinction between SUBJ and OBJ– two semantically unrestricted terms – unspecified, but these can be distinguished on the basis of the subject having a higher structural priority.

This classification is complemented by a different but related cross-classification from the Lexical Mapping Theory (LMT, Bresnan & Kanerva 1989) based on two features: $[\pm r]$ (for "(semantically) restricted") and $[\pm o]$ (for "objective"), seen in (6).

(6)

	$-r$	$+r$
$-o$	SUBJ	OBL$_\theta$
$+o$	OBJ	OBJ$_\theta$

This classification produces a markedness hierarchy of grammatical functions: SUBJ $[-r, -o] <$ OBJ $[-r, +o]$, OBL$_\theta$ $[+r, -o] <$ OBJ$_\theta$ $[+r, +o]$ (Bresnan & Moshi 1990). This hierarchy, together with the mapping principles, ensures the correct default mapping of semantic roles to grammatical functions. It also predicts the possible ways of remapping grammatical functions in passives, causatives and applicatives, although the details differ across variants (e.g. some versions of LMT allow mapping agents to OBJ and some do not). It should be stressed that LMT does not *directly* provide evidence for the set of grammatical functions, because in LFG the theory of f-structure and the theory of the mapping from semantic roles to f-structure are formally independent: one can analyze GFs without adopting any particular theory of how they are mapped to semantic roles. But indirectly, the

Oleg Belyaev

cross-classification of core GFs can serve as an independent justification for their inventory. For more information on mapping principles in LFG, see Findlay & Kibort forthcoming [this volume].

3.2 Functional hierarchy

The most fundamental distinction between grammatical functions is the universal functional hierarchy in (7), which is the LFG version of the Keenan-Comrie Hierarchy (Keenan & Comrie 1977).[3]

(7)   SUBJ > OBJ > OBJ_θ > COMP, XCOMP > OBL_θ > ADJ, XADJ

The Keenan-Comrie Hierarchy was originally devised as a typological hierarchy that constrains the range of possible grammatical functions that the relativized argument can occupy in the relative clause. It is now widely acknowledged that the same hierarchy can determine a number of grammatical processes within a single language. Phrase structure-based frameworks try to account for such generalizations by reducing the hierarchy to differences in phrase structure configuration. For example, asymmetries in anaphoric binding are typically described in terms of c-command (Chomsky 1982). In LFG, most such constraints, if they are indeed syntactic,[4] are described in terms of f-structure.[5] Thus, the relation of

---

[3]The difference from Keenan and Comrie is mainly in the terminology (OBJ_θ for what they call indirect object), but also in the split between OBJ and (X)COMP and the addition of adjuncts at the bottom of the hierarchy. Objects of comparison are not viewed as a special grammatical function in LFG and are therefore not included. Also, while Keenan and Comrie include genitive possessors, this is not done in LFG because possessors do not directly compete with clausal arguments and are somewhat special; they are discussed in Section 4.4.

[4]For many phenomena, it is not easy to decide whether the constraints should be formulated in terms of syntax, semantics, or both; in many ways this rests on the particular theories of the two and the syntax–semantics interface. For example, while mainstream generative grammar is notoriously syntactocentric, Simpler Syntax represents another extreme, where syntactic structure includes only a very basic notion of grammatical relations, and most of the work that is done by f-structure is assigned to a (very elaborate) semantic structure. As an illustration of the relationship between Culicover and Jackendoff's approach and LFG, Belyaev (2015) shows that the criteria that Culicover & Jackendoff (1997) consider to be semantic are captured at the f-structure level in LFG.

[5]It has been argued that anaphora is sometimes directly constrained by linear precedence, e.g. for Malayalam in Mohanan (1982). In LFG, this has been modeled using the f-precedence relation (Kaplan & Zaenen 1989, also see Belyaev forthcoming(a) [this volume]) by essentially stating that the c-structure nodes that map to the f-structure of the antecedent must precede the c-structure nodes that map to the f-structure of the anaphoric expression. Notably, the starting point is still the f-structure and the c-structure is only accessed through inverse mapping.

c-command is replaced by the relation of outranking in the hierarchy in (7): see Rákosi forthcoming [this volume].

3.3 Governable and ungovernable GFs

As stated above, most GFs are GOVERNABLE: that is, in LFG terms, they must appear in the list of arguments in the PRED value of their f-structure in order to be licensed. The PRED value is usually that of a verb or other clausal predicate, as in (8), which is the f-structure of the sentence *Mary ran quickly*.

$$
(8) \quad
\begin{bmatrix}
\text{PRED} & \text{'RUN\langle SUBJ\rangle'} \\
\text{TENSE} & \text{PAST} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'MARY'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\
\text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'QUICKLY'} \end{bmatrix} \right\}
\end{bmatrix}
$$

In this sentence, SUBJ is a governable GF that appears in the argument list in PRED. 'Quickly' appears as the value of the GF ADJ, which is ungovernable and is not licensed by the PRED value.

If a governable GF is included in the list of arguments in PRED but has no value, Completeness is violated; conversely, if a governable GF is present but not included in the list, Coherence is violated. Modifiers (ADJ and XADJ) are the only GFs which are ungovernable. The only condition on their occurence is that the f-structure in which they appear should have *some* PRED value.[6]

Determining exactly which GFs are arguments of a given predicate is not trivial in general, but especially in LFG because of its rigid separation between levels. Two distinctions are especially important for LFG: between semantic and syntactic argumenthood, because semantic arguments are not necessarily expressed as arguments in syntax, and vice versa (Section 3.3.1), and between arguments and adjuncts in syntax, whose status does not necessarily correlate with semantic argumenthood and adjuncthood (3.3.2).

---

[6]This constraint is part of *extended coherence* (Bresnan et al. 2016), which is not accepted by all LFG practitioners as a universal well-formedness condition. While the notion that only f-structures with PRED values can have modifiers is intuitively plausible, it is difficult to find empirical justification for this condition on adjuncts, since PRED-less f-structures normally correspond either to expletive pronouns or heads of categories like P, which both tend not to attach any modifiers at c-structure. Violation of extended coherence might be relevant for languages where some adpositions have PRED values and some do not; only the former would then be able to have adjuncts.

Oleg Belyaev

3.3.1 Semantic and syntactic arguments

First of all, one must clearly differentiate between *semantic* argumenthood and *syntactic* argumenthood. Syntactic arguments may have no semantic counterparts; such is the case of subjects of verbs like *rain*, or "raised" subjects and objects like *John* in *John seemed to come* or *David* in *I saw David come* (functionally controlled in LFG terms). In LFG, such "non-arguments" are notated as being outside the angled brackets in the argument list of the PRED feature value, e.g. 'RAIN⟨ ⟩SUBJ' or 'BELIEVE⟨SUBJ XCOMP⟩OBJ'. This effectively makes f-structure include semantic information. As discussed in Belyaev forthcoming(a) [this volume], in modern Glue Semantics-based approaches, it is possible to either completely eliminate PRED features from the syntax or at least remove semantic role information, which would make the separation between syntax and semantics more clear-cut.

Conversely, a semantic argument might have no syntactic expression. For example, unspecified object deletion or antipassivization can turn any transitive verb into an intransitive one that only has a single argument, the agent (*We ate a meal.* → *We ate.*). The semantic predicate 'eat', and the corresponding real-life event, clearly have a patient participant regardless of whether it is syntactically expressed, and this omitted participant will be interpreted in some way. But there is broad consensus in the literature (see Melchin 2019) that unspecified objects are not present in syntax in any form. In LFG, this means that they are both absent as constituents in c-structure, and as GFs in f-structure, because f-structure is a syntactic level that does not directly reflect the thematic roles of the arguments.[7]

It is important to distinguish such cases of true omission of semantic arguments at f-structure from cases where arguments do not overtly appear at c-structure but are still present at f-structure. Two widespread cases when this occurs are pro-drop (like in Italian *ha vinto* lit. 'has won' = 's/he has won') and raising (*John seems to have won*, where *to have won* appears to lack a subject). The "little *pro*" analysis of null subjects in languages like Italian has been assumed at least since Perlmutter (1971) and is supported by much empirical evidence, such as the possibility of controlling PRO, serving as the antecedent of anaphors, controlling agreement etc. that is well-known from basic syntax textbooks and need

---

[7]The mapping from semantic roles to GFs is handled in LFG by a separate component, Linking Theory. In the most widespread variant of Linking Theory, Lexical Mapping Theory, unspecified object deletion is captured by suppressing the realization of the patient argument, i.e. preventing it from being mapped to any GF. See Findlay & Kibort forthcoming [this volume] for further explanation.

not be repeated here. This evidence is also valid in LFG and leads one to conclude that while *pro* is not needed at c-structure, it has to be present at f-structure in subject position. Similarly, "raised" (functionally controlled) arguments overtly appear in main clauses but still have to satisfy the subcategorization constraints of the embedded clause. In the LFG analysis of raising, one f-structure is shared between the main clause subject or object and the subject of the embedded clause. Therefore, both components of the functional control relation are present in the syntax as arguments of their respective clauses; see Section 4.3.2 and Vincent forthcoming [this volume] for more detail.

3.3.2 Arguments and adjuncts

In one form or another, the problem of arguments vs. adjuncts is relevant for all grammatical frameworks, but LFG is special in that it treats the syntactic distinction between arguments and adjuncts as fully separate from the homonymous semantic distinction. The syntactic distinction between arguments and adjuncts also does not exist in other frameworks in the same form; for example, the HPSG approach is typically to include all verbal dependents in an ordered list DEPS. This means that semantic subcategorization and semantic obligatoriness cannot be used as a reliable criteria by themselves: it was shown above that semantic arguments might not correspond to any GF in syntax. Similarly, some analyses treat passive agents as adjuncts, in spite of their semantic argumenthood. The issue is further complicated by the fact that additional, derived arguments that are not present in the lexical entry of the predicate can be introduced in the syntax (Needham & Toivonen 2011). Hence, criteria for distinguishing between arguments and adjuncts must be purely syntactic.

The main empirical difference between arguments and adjuncts can be formulated in terms of Dowty's (1982) *subcategorization test*: Modifiers, but not arguments, can be omitted. In a theory like LFG which uses no empty heads (see Belyaev forthcoming(a) [this volume]), this criterion is clearly not general enough, because grammatical functions that are present at f-structure may lack a realization at c-structure, e.g. under pro-drop (see Section 3.3.1 above). Normally, the presence of such "null" elements like *pro* and their features is reflected in the morphology through agreement or argument incorporation, although some languages, like Japanese, are notorious for allowing almost unrestricted pro drop — for these languages, distinguishing between arguments and adjuncts using the subcategorization test is especially problematic.

Another truly syntactic criterion is that adjuncts can be freely multiplied in any number, whereas arguments cannot (Kaplan & Bresnan 1982: 40):

Oleg Belyaev

(9)   The girl handed the baby a toy **on Tuesday**$_{ADJ}$ **in the morning**$_{ADJ}$.

(10)   * The girl saw **the baby**$_{OBJ}$ **the boy**$_{OBJ}$.

Crucially, the multiplication test is only relevant for adjuncts *of the same type*. While a clause may have at most one subject and object, it may have several obliques or indirect objects (as elaborated in sections 3.5.2 and 3.5.3 below). But there can still be only one indirect object or oblique with the same semantic role:

(11)   * John went to Moscow to Red Square.

Other criteria have to do with the specific understanding of grammatical functions in LFG, their relative ordering and the licensing of long-distance dependencies. For example, some pronouns, such as the reflexive pronoun *seg selv* in Norwegian, are specifically limited in their coreference to coarguments (Hellan 1988), and therefore cannot occur in adjunct position. The examples in (12) are cited from Dalrymple et al. (2019). In (12a) this reflexive is a direct object that is coreferent to the subject – both are arguments. Similarly, in (12b), the reflexive is used in a PP that is an oblique argument selected by the verb 'tell'. But in (12c), the prepositional phrase containing the reflexive is not an argument of the predicate and thus it cannot have the subject as its antecedent. Thus the cut-off point in the hierarchy in (7) for *seg selv* is just to the right of ADJ, XADJ.

(12)   Norwegian (Germanic > Indo-European)

    a.   Jon forakter **seg selv**.
       Jon despises self
       'Jon$_i$ despises **himself**$_i$.'

    b.   Jon fortalte meg **om**   **seg selv**.
       Jon told   me   about self
       'Jon$_i$ told me **about himself**$_i$.'

    c.   * Hun kastet meg **fra**   **seg selv**.
        she   threw me   from self
        'She$_i$ threw me away **from herself**$_i$.'

It is also widely assumed in the literature that wh-extraction from adjuncts is impossible (Pollard & Sag 1987, Huang 1982, Rizzi 1990). However, this constraint does not seem to be cross-linguistically universal, or at least it does not apply to all types of modifiers. For example, while extraction from clausal adjuncts is prohibited (13), extraction from PPs is allowed (14).

(13)   * **Which man** did John leave when he saw __?

(14)   **Which bed** did David sleep **in** __?

3.4 Terms and non-terms

Another distinction is between core arguments, or terms, and non-core arguments, or non-terms.

(15) $\underbrace{\text{SUBJ} > \text{OBJ} > \text{OBJ}_\theta}_{\text{terms}} > \text{COMP, XCOMP} > \text{OBL}_\theta > \text{ADJ, XADJ}$

There is no universal set of tests that distinguishes between terms and non-terms, but a number of constructions in different languages are systematically sensitive to this distinction; see Alsina (1993) for a detailed discussion of termhood. Some of these constraints are discussed in the following sections.

3.4.1 Agreement

In many languages, verb agreement seems to be only possible with terms, that is, subjects, objects or secondary objects. The idea goes back at least to Johnson (1977: 157), where it is called the Agreement Law. It has the same status in Relational Grammar (Frantz 1981). Agreement with subjects is very widespread; many languages also have (obligatory or optional) agreement with direct objects; the map in WALS (Siewerska 2013) cites 193 such languages among the sample of 378, i.e. more than 50%. Indexing other arguments is considerably less common, but some languages also agree with secondary objects, like Basque (isolate). As seen in (16), finite ditransitive verbs in Basque agree with the ergative (SUBJ), absolutive (OBJ) and dative (OBJ$_\theta$) arguments in person and number.

(16) Basque (isolate)
      d-a-kar-ki-da-zu
      3SG.ABS-PRS-bring-DAT-1SG.DAT-2SG.ERG
      'you bring it to me' (Hualde et al. 2003: 209)

From current LFG literature, it is unclear whether the restriction of agreement to terms is a theoretical postulate or an empirical observation, since the termhood of agreement controllers is usually confirmed by independent syntactic evidence.

3.4.2 Control

Cross-linguistically, only terms tend to be controllers or controllees in control constructions, both lexically determined (clausal complements) and not (clausal adjuncts). This is certainly true for English, where only subjects and direct objects can control null PRO in control constructions. In French, controllers may,

Oleg Belyaev

in addition to subjects (17) and objects (18), also be secondary objects introduced by the preposition *à* (or expressed by dative clitics), as in (19).[8]

(17) French (Romance > Indo-European)
Jean a promis [de venir demain].
    has promised to come tomorrow

(18) Jean a convaincu **Marc** [de venir demain].
    has convinced M. to come tomorrow

(19) Jean a demandé **à Marc** [de venir demain].
    has asked to M. to come tomorrow

    Similarly, Kroeger (1993) shows that in Tagalog, only terms can be controllees in participial complement constructions and controllers in adjunct constructions. According to Kroeger, voice suffixes in Tagalog promote any argument to subject status, and the erstwhile subject (the agent) becomes an OBJ$_\theta$ (see an illustration in (26) below).[9] Thus, (20) illustrates the verb 'read' in the active voice; the controller is the subject. In (21), the verb 'read' is marked by the "object voice" suffix: the Patient is promoted to subject status and carries the nominative proclitic *ang=*. The controllee is still the Agent, which in this example is demoted to OBJ$_{AG}$. Finally, (22) shows that obliques, i.e. arguments that are not subjects, direct objects or demoted agent-like arguments in voice constructions, cannot be controllees, even if they have the same semantic role Agent.

(20) Tagalog (Malayo-Polynesian > Austronesian)
In-abut-an ko siya=ng [nagbabasa ___SUBJ ng=komiks
PFV-find-DV 1SG.GEN 3SG.NOM=COMP AV.IPFV.read GEN=comics
sa=eskwela].
DAT=school
'I caught him reading a comic book in school.' (Kroeger 1993, via Dalrymple et al. 2019: 16)

---

[8] An anonymous reviewer wonders whether French PPs with the preposition *à* should be viewed as OBL$_\theta$ rather than OBJ$_\theta$, like their English counterparts with *to*. It is commonly assumed that in French these are, indeed, secondary objects, cf. Manning (1992), Yates (2002), Lovestrand (2009), although some authors treat them as OBL$_\theta$, e.g. Schwarze (2001). In my view, the control facts cited below do suggest that these PPs are secondary objects rather than obliques.

[9] Such arguments must be treated as secondary objects because they are marked by the same genitive proclitic *ng=* that marks direct objects, which do not change their mapping when an agent is demoted.

(21) In-iwan-an   ko   siya=ng   [sinususulat   __OBJ_AG ang=liham].
PFV-leave-DV 1SG.GEN 3SG.NOM=COMP IPFV.write.OV   NOM=letter
'I left him writing the letter.' (Kroeger 1993, via Dalrymple et al. 2019: 16)

(22) *In-abut-an ko   si=Luz   na [ibinigay   ni=Juan
PFV-find-DV 1SG.GEN NOM=Luz LINK IV.IPFV.give GEN=Juan
ang=pera   __OBL_GOAL].
NOM=money
('I caught Luz being given money by Juan.') (Kroeger 1993, via Dalrymple et al. 2019: 16)

3.4.3 Reflexivization

Kibrik (2000) argues that in Archi (Lezgic > East Caucasian), most constructions do not single out any argument as the grammatical "subject" or "object". When there is any difference at all, it is between core arguments (terms) and non-core arguments (non-terms). Kibrik shows that any of the core arguments (subject or direct object) can be the controllee in control constructions, similarly to the Tagalog data above. But Kibrik also claims that possessive reflexives can be controlled by the subject or the object, in any direction (23a), but not by non-core-arguments (24).

(23) Archi (Lezgic > East Caucasian)

a. tow-mu_i žu-n-a-ru_i   ɬ:onnol a<r>č-u
   he-ERG   self-GEN-EMPH-CL2 wife(CL2) ⟨CL2⟩kill-PRF
   'He_i (pron., erg.) killed his (refl.) wife_i (abs.).' (A > P)

b. tor_i že-n-a-w_i   bošor-mu   a<r>č-u
   she self-GEN-EMPH-CL1 husband(CL1)-ERG ⟨CL2⟩kill-PRF
   'Her_i (refl.) husband (erg.) killed her (pron., abs.).' (P > A)

(Kibrik 2000: 62)

(24) a. tow_i žu-n-a-bu_i   abej.me-t:i-š   k'olma-ši
      he   self-GEN-EMPH-CL1.PL parents(CL1.PL)-SUPER-EL separate-ADV
      w-i
      CL1-be.PRS
      'He_i (pron., abs.) lives apart from his_i (refl.) parents.' (SUBJ > OBL)
      (ibid.)

Oleg Belyaev

    b. *tow.mu-t:i-š$_i$ žu-n-a-bu   abaj   k'olma-ši
       he-SUPER-EL self-GEN-EMPH-CL1.PL parents(CL1.PL) separate-ADV
       b-i
       CL1-be.PRS
       ('His$_i$ (refl.) parents (abs.) live apart from him$_i$ (pron., abs.).') (OBL >
       ABS) (ibid.)

Therefore, while subject-oriented reflexives are found in many languages (see
Dalrymple 1993 and Rákosi forthcoming [this volume]), Archi is different in hav-
ing subject *and* object, i.e. term-oriented, reflexives.

3.5 Semantically restricted and unrestricted arguments

The classification of GFs into terms and non-terms allows us to distinguish be-
tween subjects, objects and all other grammatical functions. But the difference
between "primary" and "secondary" objects (OBJ and OBJ$_\theta$) remains undefined.
This distinction is captured by another classification of GFs into semantically
restricted and unrestricted arguments:

(25) $\underbrace{\text{SUBJ} > \text{OBJ}}_{\text{semantically unrestricted}}$ $> \text{OBJ}_\theta > \text{COMP, XCOMP} > \text{OBL}_\theta > \text{ADJ, XADJ}$

As mentioned above, θ in the GF names OBJ$_\theta$ and OBL$_\theta$ stands for a particular
thematic role that is filled by this argument. Thus they are families of GFs, each
of which is associated with a particular semantic role: OBL$_{GOAL}$, OBJ$_{THEME}$, etc. In
this, they are contrasted with subjects (SUBJ) and direct objects (OBJ), which do
not have this additional qualifier attached to them.

    The specific list of thematic roles is not agreed upon in LFG. In the case of OBJ$_\theta$
and OBL$_\theta$, it is not even clear whether the roles that appear in θ are universal or
language-specific (the fact that θ is often equivalent to the PCASE value supplied
by an adposition suggests the latter). For more information on the mapping from
thematic roles to GFs, see Findlay & Kibort forthcoming [this volume].

    A consequence of the distinction between semantically restricted and seman-
tically unrestricted GFs is the fact that only the latter can be non-arguments at
the semantic level; the former must be assigned some thematic role. This, in turn,
predicts that, first, arguments lacking any semantic role (expletives or dummy ar-
guments) like English *it* or *there* (such as in *It rained*) can only appear in subject
or direct object position; second, that "raising" (functional control in LFG terms)
is only possible when the matrix clause position is SUBJ or OBJ.

    In what follows, I will discuss the motivation for treating each of these GFs as
semantically restricted or unrestricted in detail.

3.5.1 Unrestricted GFs

3.5.1.1 Subjects   One of the key features of subjects is that they are not restricted to one semantic role (Fillmore 1968). The semantic unrestrictedness of subjects is perfectly illustrated by the existence of passive constructions: the same lexical verb can have either the Agent (in the active voice) or the Patient (in the passive voice) as its subject. Some languages go even further and allow promoting any argument to subject status if it has discourse prominence, or for syntactic reasons. One such language is Tagalog, where the voice suffix on the verb determines which argument bears the SUBJ GF, according to the analysis in Kroeger (1993):

(26)   Tagalog (Greater Central Philippine > Austronesian)

    a.   active voice
       B-um-ili   **ang=lalake** ng=isda   sa=tindahan.
       PRF.AV-buy NOM=man   GEN=fish DAT=store
       '**The man** bought fish at the store.'

    b.   objective voice
       B-in-ili-∅   ng=lalake **ang=isda** sa=tindahan.
       PRF-buy-OV GEN=man NOM=fish DAT=store
       'The man bought **the fish** at the store.'

    c.   dative voice
       B-in-ilh-an ng=lalake ng=isda   **ang=tindahan**.
       PRF-buy-DV GEN=man GEN=fish NOM=store
       'The man bought the fish **at the store**.'

    d.   instrumental voice
       Ip-in-am-bili ng=lalake ng=isda   **ang=pera**.
       IV-PRF-buy   GEN=man GEN=fish NOM=money
       'The man bought fish **with the money**.'

    e.   benefactive voice
       I-b-in-ili   ng=lalake ng=isda   **ang=bata**.
       BV-PRF-buy GEN=man GEN=fish NOM=child
       'The man bought fish **for the child**.'

The formal marking of the subject is also not usually directly derived from its semantic role. We saw above that in Tagalog, the subject always receives the nominative preposition *ang*. In languages where non-canonical subject marking is possible, there is still no consistent association between case marking and the

semantic role of the subject. For example, Icelandic oblique subjects are never agent-like, but the choice of the case marker does not otherwise consistently correlate with particular semantic roles (Jónsson 2003). Even among Daghestanian (East Caucasian) languages, where experiencer subjects are regularly marked by dative instead of ergative, there is some variation as to which case is selected by which verb; for example, in Gubden Dargwa, the verb 'see' selects ergative case and the verb 'want' selects dative case, while in the closely related Khuduts Dargwa both verbs have dative subjects (Ganenkov 2013: 246).[10] In short, subjects are usually consistently encoded regardless of their semantic role, and when there is variation in marking, it is usually lexical and idiosyncratic.

3.5.1.2 Objects   Objects, too, are not associated with specific semantic roles. While objects are never agents in English, they can still have a range of semantic roles: Patient (*John ate **the cookie***), Stimulus (*John saw **David***), Experiencer (*It surprised **me***), Theme (I gave **the book** to John). Just like Tagalog can promote various arguments to subjects, some languages allow promoting arguments to direct objects via so-called applicative constructions. One such language is Hakha Lai (Tibeto-Burman > Sino-Tibetan), which I describe following Peterson (2007: 15ff.). In Hakha Lai, verbs agree with two core arguments – subjects and objects – of transitive verbs, as in (27).

(27)   Hakha Lai (Tibeto-Burman > Sino-Tibetan)
    ʔan-kan-thoʔŋ
    SUBJ.3PL-OBJ.1PL-hit
    'They hit us.' Peterson (2007: 16)

It can be reasonably assumed that, in LFG terms, the argument indexed by the first prefix is SUBJ, while the argument indexed by the second prefix is OBJ.

Hakha Lai also has a range of applicative suffixes that introduce additional morphologically unmarked arguments into the verb's argument structure. One such marker is the benefactive / malefactive suffix *-piak*. When this suffix is used, it is the newly introduced argument that occupies the OBJ position, as seen from the agreement pattern in (28). The verb agrees with the first person singular benefactive argument ('on me') and not with the third person singular patient ('wood slab').

---

[10]It is worth mentioning that it is not clear whether Daghestanian languages have the GF SUBJ. Kibrik (2000) argued that only core arguments (terms in LFG) can be distinguished, but there is no evidence for the privileged status of either of the core arguments. The universality of subjects is discussed in Section 4.2.4.

(28) ʔaa!   tleem-pii   khaa maʔ-tii tsun taar-nuu=niʔ?
      INTERJ wood.slab-AUG DEIC DEM-do DEIC old-woman=ERG
      ʔa-**ka**-khaʔŋ-piak=ʔii...
      SUBJ.3SG-OBJ.1SG-burn-MAL=CONN
      'Ah, the old woman burned the big slab of wood on me, and...'

3.5.2 Obliques

The reason for treating obliques as semantically restricted and a family of functions is that, unlike subjects and objects, their marking will always vary depending on their semantic role. For example, Goals in English use the preposition *to* (as in *Mary went to London*), while Sources use the preposition *from* (*David came from Paris*). This justifies treating OBL as a family of functions rather than a single GF.

Another reason for this architectural decision is that there may be multiple obliques in one clause. In English, this can be illustrated by sentences like *John moved from London to Paris*, where *from London* can be analyzed as OBL$_{SOURCE}$ and *to Paris* as OBL$_{GOAL}$. This can be disputed, however, because either of the obliques, or both, can be omitted; thus Zaenen & Crouch (2009) propose doing away with OBL together, replacing OBL with set-valued ADJ. In other languages, however, the evidence for multiple OBL arguments can be more compelling. Dahlstrom (2014) shows that in the Algonquian language Meskwaki, obliques are strictly positioned immediately before the verb (29), while other arguments (subjects, objects, secondary objects and complement clauses) appear postverbally, as illustrated in (30), where 'Wisahkeha' is analyzed as a direct object by Dahlstrom.

(29) Meskwaki (Algonquian > Algic)
      aˑkwi **nekotahi** wiˑhnahi-ihaˑyanini
      not   anywhere FUT-be.in.habit.of-go(thither)-2/NEG
      'You will never go **anywhere**.' (Dahlstrom 2014: 57)

(30) iˑni=keˑhi=ipi=meko   eˑh-awataw-aˑči   wiˑsahkeˑh-ani
      then=and=HRSY=EMPH AOR-take.OBJ2.to-3>3′/AOR W.-SG
      metemoˑh-eˑh-a
      old.woman-DIM-SG
      'And right then, it's said, the old woman took it to Wisahkeha.'
      (Dahlstrom 2014: 58)

In Meskwaki, obliques are not optional but required by verbal stems or preverbs. For example, all verbs of quotation require an oblique argument. Therefore,

the participants that Dahlstrom terms "obliques" cannot be analyzed as adjuncts in terms of LFG.

Dahlstrom further demonstrates that additional oblique arguments may be associated with preverbs (which can be viewed as a kind of applicative marker) or compounded verb stems. When more than one oblique appears in a clause, all must precede the verb, and each oblique argument must be adjacent to the associated root or preverb.

(31) awitameko   **keˑkoˑhi** iši–   **ateškaˑwi**   –išawihkapa
      not.POT=EMPH any.way thus– with.delays –thus.happen.to.S-2/POT
      'You would not have experienced delays in any way.' (Dahlstrom 2014: 64)

In (31), *ateškaˑwi* is associated with the verb stem, while *keˑkoˑhi* satisfies the valency introduced by the preverb *iši-*. The special position of each of these obliques seems to present compelling evidence for treating them as separate (though related) semantically restricted GFs.

3.5.3 Secondary objects

Among all the main GFs, secondary objects are perhaps the most difficult to characterize. They are similar to objects in being terms, and to obliques in being semantically restricted. But these classifications are not easily translatable into specific empirical properties. We have seen above that secondary objects are similar to direct objects in being terms, which allows them to control verbal agreement and PRO. But these criteria do not always serve to distinguish OBJ_θ; for example, neither applies to English. Another property of secondary objects, which likens them to obliques, is their semantic restrictedness.

Secondary objects were originally thought of as occupying a single GF OBJ2 and identified on the basis of constructions like (32) in English and other Germanic languages like Icelandic. In English, the identification of OBJ2 is straightforward due to the fact that it is the only argument apart from subject and direct object that is not marked by a preposition (which is a feature of obliques) and also due to the alternation of the double object construction in (32) with the oblique dative construction in (33). In the latter, the same thematic roles map to two constructions that differ both in word order and case / preposition marking. This means that at least three different GFs must be distinguished: OBJ, OBL_θ and OBJ_θ.

(32) John gave [Mary]$_\text{OBJ1}$ [a book]$_\text{OBJ2}$.

(33) John gave [a book]$_\text{OBJ1}$ [to Mary]$_\text{OBL}$.

The fact that *Mary* is indeed the direct object in (32), even though it is called an "indirect object" in traditional grammar (due to its dative semantics), can be shown from the fact that in the passive version of (32), it is the recipient that is promoted to subject status (34).

(34)   a. Mary was given a book.
       b. * A book was given Mary.

Passivization is not a *direct* criterion for objecthood, because in LFG the passive is a lexical process and not a syntactic transformation (Dalrymple et al. 2019: 28). But indirectly, lexical mapping constraints do determine which arguments can be passivized. Objects can be passivized because they are inherently classified as $[-r]$, and, in the absence of a higher-ranking argument, fill the SUBJ GF which is defined as $[-o, -r]$. Secondary objects, in contrast, cannot be passivized because they are inherently defined as $[+o]$. This is one of the key features of secondary objects as opposed to direct objects.

In English, the label OBJ2 may indeed be appropriate, because there can be only one secondary object, and this object is connected to only one semantic role (Theme). But other languages make much wider use of secondary object functions, such that there may be several OBJ$_\theta$s, each of which is restricted to a different semantic role. For example, Bresnan & Moshi (1990) analyze Kichaga (Bantu) as having verbal indexing of multiple thematically restricted objects, each of which has its own slot in the verb form:

(35)   Kichaga (Bantu)
       n-ä-lʲé-kú-shí-kí-kóṛ-í-à
       FOCUS-1SUBJ-PST-17OBJ-8OBJ-7OBJ-cook-APPL-FV
       'She/he cooked it with them there.' (Bresnan & Moshi 1990)

Of the three object prefixes in this example, only the instrumental object (8OBJ-) is unrestricted; the other two are thematically objects that occupy the grammatical functions OBJ$_{LOC}$ (17OBJ-) and OBJ$_{PATIENT}$ (7OBJ-).

Another use of OBJ2 / OBJ$_\theta$ is to capture the difference between case-marked (topical) and unmarked objects in languages with Differential Object Marking (DOM), where the direct object can either be marked by a special (accusative) case or left unmarked (as discussed in Section 2; also see Butt forthcoming [this volume]). According to Dalrymple & Nikolaeva (2011), in many such systems, accusative-marked direct objects have the GF OBJ, while unmarked objects are OBJ$_\theta$. The same distinction may be reflected in agreement patterns: Dalrymple

Oleg Belyaev

& Nikolaeva (2011) show that in Ostyak (Ob-Ugric > Uralic), objects that trigger agreement are OBJ while objects that do not are OBJ_θ. With respect to case marking, an opposite viewpoint is taken by Butt & King (1996), who treat focal, unmarked objects as OBJ. It may be that different patterns are observed in different languages. It is also possible that in some languages, the distinction is not reflected by any overt case marking or agreement; the theory itself does not constrain this in any way.

3.5.4 Universality

From these examples it is clear that secondary objects are very similar to obliques in being semantically restricted and covering a similar set of semantic roles. Secondary objects have to be recognized only in those languages where there is evidence that some arguments are more prominent than obliques (e.g. in case marking, verb morphology, or anaphora) but less prominent than direct objects. Not all grammars involve such fine-grained distinctions, and in this sense OBJ_θ is probably not universal.

In contrast, OBL_θ as it is understood and used in LFG is, in effect, architecturally necessary,[11] because SUBJ and OBJ provide only two positions, which is not enough to map all possible thematic roles that verbs may have.

Finally, it is theoretically possible that some languages do not make use of the GF OBJ. Such a language would have only one semantically unrestricted function, SUBJ; all other arguments would be OBJ_θs or OBL_θs with various semantic roles. It would also lack a passive, because, under Lexical Mapping Theory, passivization depends on the presence of a second $[-r]$ argument that is promoted to subject status. In effect, this would be a language where most semantic roles are directly encoded in the syntax, i.e. there is a one-to-one correspondence between GFs and thematic roles, except for one unrestricted argument. This idea has been discussed in two distinct flavours. Börjars & Vincent (2008) consider whether the OBJ vs. OBJ_θ distinction should be abandoned as such (i.e. all objects in all languages are OBJ_θs). In contrast, Lander et al. (2021) make this proposal for the specific case of West Circassian (West Caucasian). West Circassian, a polysynthetic language, has a rather unusual system of applicative prefixes that is unlike the more typologically common system discussed above for Hakha Lai: see (28) above. In Hakha Lai, additional arguments introduced by applicative morphology are promoted to OBJ status, while the erstwhile object is demoted to OBJ_θ.

---

[11]Assuming that OBJ_θ is not universal. Logically speaking, if the language only draws a distinction between SUBJ, OBJ, and all other arguments, it does not matter whether the latter are called OBJ_θ or OBL_θ.

In West Circassian, applicative prefixes simply introduce additional arguments without altering the status of existing arguments. The absolutive argument is not indexed on the verb and the corresponding full NP (if present) bears Absolutive case. All other arguments are introduced by prefixes and their full NP counterparts bear Oblique case. For example, in (36) the Absolutive Patient is 'dishes' (*laʁe-xe-r*) and has no corresponding verbal prefix. The three other arguments bear Oblique case: 'boy' (*č'ale-m*) corresponds to the prefix *jə-*, 'girl' (*pŝaŝe-m*) corresponds to ∅-*r-* and 'you' is expressed only by the prefix *b-də-*.

(36) West Circassian (West Caucasian)
    č'ale-m$_i$ pŝaŝe-m$_j$ laʁe-xe-r   we$_k$
    boy-OBL girl-OBL dish-PL-ABS you.SG
    qə-b-$_k$də-∅-$_j$r-jə-$_i$tə-ʁe-x
    DIR-2SG.IO-COM-3SG.IO-DAT-3SG.ERG-give-PST-PL
    'The boy gave the dishes to the girl with you (sg.).'

Lander et al. (2021) argue for a syntactically ergative analysis of West Circassian, showing that the Absolutive argument has privileged status in certain constructions; it is assigned the grammatical function SUBJ. In contrast, they find no evidence for a distinction between different types of indexed arguments and analyze them all as OBJ$_\theta$: ergative agents are OBJ$_{AGENT}$, recipients are OBJ$_{RECIP}$, instrumentals are OBJ$_{INSTR}$ etc. Thus the sentence (36) gets the f-structure (37) in their analysis.

$$(37) \begin{bmatrix} \text{PRED} & \text{`GIVE}\langle\text{SUBJ, OBJ}_{AG}\text{ , OBJ}_{GOAL}\text{ , OBJ}_{COM}\rangle\text{'} \\ \text{TENSE} & \text{PAST} \\ \text{DIR} & Q\partial \\ \text{SUBJ} & \begin{bmatrix} \text{PRED `DISH'} \\ \text{PERS 3} \\ \text{NUM PL} \end{bmatrix} \\ \text{OBJ}_{AG} & \begin{bmatrix} \text{PRED `BOY'} \\ \text{PERS 3} \\ \text{NUM SG} \end{bmatrix}_i \\ \text{OBJ}_{GOAL} & \begin{bmatrix} \text{PRED `GIRL'} \\ \text{PERS 3} \\ \text{NUM SG} \end{bmatrix}_j \\ \text{OBJ}_{COM} & \begin{bmatrix} \text{PRED `PRO'} \\ \text{PERS 2} \\ \text{NUM SG} \end{bmatrix}_k \end{bmatrix}$$

Oleg Belyaev

According to Lander et al., West Circassian does make use of the grammatical function OBL$_\theta$ for those arguments that are not indexed and are marked by postpositions, but there is no need for the grammatical function OBJ in this language.

4 Individual GFs

4.1 General remarks

In the preceding section, I described the cross-classification of grammatical functions according to three parameters: governability, termhood and semantic restrictedness. This subdivides the main GFs into four classes: (1) SUBJ and OBJ (governable semantically unrestricted terms); (2) OBJ$_\theta$ (governable semantically restricted term); (3) OBL$_\theta$ (governable semantically unrestricted non-term); (4) ADJ (ungovernable). However, this is not enough to characterize all the grammatical functions for the following reasons. First, (x)COMP and POSS, being restricted to rather specific syntactic configurations, do not readily fit into this picture: while (x)COMP is clearly governable, it is not clear whether it is a term; as for POSS, while it is clearly semantically unrestricted, it is not clear whether it is a term and whether it is, in fact, governable. Secondly, the distinction between SUBJ and OBJ remains unspecified.[12] Thirdly, the cross-classification of grammatical functions is not meant to explain all of their properties: even grammatical functions like OBJ$_\theta$, whose existence is predicted by the cross-classification itself, may have individual properties that do not follow from their class membership.

Therefore, in this section, I will proceed from the "big picture" drawn above towards characterizing the unique properties of each of the standard grammatical functions in LFG, sometimes together with other GFs in order to provide a better contrast. Subjects are opposed to all other grammatical functions and will be discussed separately in Section 4.2. Many LFG approaches treat clausal complementation and nonverbal predication similarly, and both are discussed in Section 4.3. The treatment of possessors in LFG is rather special: in many ways they are like subjects, but they are also sometimes viewed as being ungovernable, likening them to adjuncts instead. Accordingly, they are given a separate treatment in Section 4.4.

---

[12]As mentioned above, Lexical Mapping Theory classifies them both as semantically unrestricted [−r], but OBJ is "objective" [+o] while SUBJ is not [−o]. But this distinction only plays a role in mapping thematic roles to grammatical functions; it is not relevant for the actual syntactic properties of subjects and objects, which is the focus of this chapter.

4.2 Subjects

All grammatical frameworks that have any notion of grammatical function assign a special status to the subject. Its properties are mainly derived from its position at the top of the functional hierarchy, and are discussed in Section 4.2.1. The centrality of the subject also raises the question of its universality, which can be approached from three different perspectives. First, is the subject universal across sentences within a single language, i.e. do all sentences have to have a subject (Section 4.2.2)? Secondly, do all languages map semantic arguments to subjects in the same way? For example, do ergative languages employ the same mapping as accusative languages? This is discussed in Section 4.2.3. Finally, is the notion of subject universal at all – are there languages where no single argument can be identified as the priority target of most syntactic relations and processes (Section 4.2.4)?

4.2.1 Core properties

The subject can be characterized as the most prominent argument in the clause, both in terms of the hierarchy in (7) and in that it is usually the sentence topic (at least in syntactically accusative languages). As with all other GFs, there is no specific set of tests that would define subjects cross-linguistically. Rather, being highest-ranking in the Keenan-Comrie hierarchy, they are expected to always participate in processes that are dependent on this hierarchy. More specifically, if a syntactic construction always targets only one argument of a clause, this argument is likely to be identified as the subject. Many tests for subjecthood have been proposed in the literature (for one summary, see Andrews 2007);[13] in the end, the particular set of diagnostics should be identified on a language-by-language basis.

One diagnostic is agreement. We have seen above that cross-linguistically, only terms can control agreement. But if any one term is the sole agreement controller in a language, this has to be the subject. Moravcsik (1978: 364) proposes a typological universal: If a language has agreement with anything other than an intransitive subject, it also has to exhibit agreement with the intransitive subject. Note that this universal is carefully formulated to include ergative

---

[13]Subject criteria that are commonly proposed in the literature include: case marking and agreement; ellipsis under coordination; binding of reflexive pronouns; control of null subjects (PRO) of infinitives and gerunds; selection in switch reference systems (same-subject / different-subject). Many more language-specific tests have been proposed as well.

Oleg Belyaev

languages (which only show S/P agreement) and does not automatically iden-
tify the subject in the "accusative" sense (A/P). I will return to the question of
subjecthood in non-accusative languages below.

It also seems that only subjects can be "raised", i.e. in LFG terms, shared (func-
tionally controlled) with a term argument in the main clause. English only has
subject-to-subject (38) and subject-to-object (39) raising.

(38)   **John** seemed [__ to agree].

(39)   John believed **David** [__ to be crossing the street].

In Icelandic, the raising rule also applies to non-nominative ("quirky") subjects
(Andrews 1982). Thus, in (40a–c) the verbs select accusative, dative and genitive
subjects, respectively.

(40)   Icelandic (Germanic > Indo-European)

    a.   **Drengina**   vantar mat.
        boys.DEF.ACC lacks   food.ACC
        '**The boys** lack food.'

    b.   **Barninu**   batnaði   veikin.
        child.DEF.DAT recovered.from disease.DEF.NOM
        '**The child** recovered from the disease.'

    c.   **Verkjanna**   gætir   ekki.
        pains.DEF.GEN is.noticeable not
        '**The pains** are not noticeable.'

This case marking is retained under raising in the main clause (41). These exam-
ples also illustrate how subjecthood is independent not only from semantic role,
but also from case marking.

(41)   a.   Hann telur   **mig**   (í barnaskap sínum) [vanta peninga].
        he   believes me.ACC in foolishness his   to.lack money.ACC
        'He believes **me** (in his foolishness) to lack money.'

    b.   Hann telur   **barninu**   (í barnaskap sínum) [hafa
        he   believes child.DEF.DAT in foolishness his   to.have
        batnað   veikin].
        recovered.from disease.DEF.NOM
        'He believes **the child** (in his foolishness) to have recovered from the
        disease.'

c. Hann telur   **verkjanna**   (í barnaskap sínum) [ekki gæta].
   he   believes pains.DEF.GEN in foolishness his   not   noticeable
   'He believes **the pains** (in his foolishness) not to be noticeable.'

4.2.2 Subjectless sentences?

A persistent question in theoretical linguistics is whether subjects are universal — that is, if subjectless sentences exist. Note that the very fact that this is a valid question follows from the LFG assumption that GFs like SUBJ are theoretical primitives (even if they tend to be associated with a set of typical empirical diagnostics). Were the subject only defined as the highest-ranking argument in a list of ARGS (as in Simpler Syntax and some variants of HPSG), each clause would automatically have a "subject" as long as its predicate had any syntactic arguments. In LFG, subjects are also assumed to be, by and large, prominent in different senses (more on this below), but this does not entail that subjectless sentences cannot exist, if only at the periphery of grammar.

That being said, the Subject Condition in (42) is widely assumed to hold in LFG (Bresnan & Kanerva 1989) — as a theoretical stipulation, not as a consequence of the framework's architecture. Most versions of the Lexical Mapping Theory also predict that one of the arguments will always be mapped to SUBJ.

(42)   Subject Condition:
       Every verbal predicate must have a subject.

The Subject Condition certainly holds in English, as well as in many other languages. But is it universal? Examples like (43) from German and (44) from Russian at first sight seem to be exceptions to the Subject Condition.

(43)   German (Germanic > Indo-European)
       … weil   getanzt wird
         because danced become.PRS.3SG
       'because there is dancing'

(44)   Russian (Slavic > Indo-European)
       menja tošnit
       I.ACC nauseate.PRS.3SG
       'I feel sick.'

The German example in (43) has an intransitive verb with no overt arguments, even though German is generally not a pro-drop language. The Russian verb in

Oleg Belyaev

(44) only has an accusative experiencer argument; while Russian does allow null subjects, it does so in a limited number of contexts and always optionally, while here no nominative argument can be expressed. However, Berman (1999, 2003) argues that the agreement morphology indicates that German examples contain a null expletive subject with only PERS and NUM features, and no PRED value. The same analysis can be extended to the Russian data.

A more convincing case for subjectless sentences is found in the Polish examples like (45), discussed in Kibort (2006). In this construction, the verb stands in the infinitive form, thus having no agreement morphology. To Kibort, this indicates that such sentences are truly subjectless. The agent may be optionally expressed, but as an oblique prepositional argument — not as a subject.

(45) Polish (Slavic > Indo-European)
Słychać ją / jakieś mruczenie.
hear.INF her.ACC some.N.ACC murmuring(N).ACC
'One can hear her/some murmuring.'

Subjectless sentences also appear in Lowe et al.'s (2021) analysis of the Sanskrit raising verb *śak* 'can'. When this verb is passivized, one of the possible outcomes is for the raised subject of the subordinate clause to stand in the instrumental case, while the object remains in the accusative:

(46) Sanskrit (Indo-Aryan > Indo-European)
rājabhī rāmaṃ hantuṃ na śakyate
kings.INS R.ACC slay.INF not can.PASS.3SG
'Rāma cannot be slain by the kings.'

Lowe et al. argue that in this construction the matrix clause has two arguments: the instrumental as OBL$_\theta$ and the subordinate clause as XCOMP, and thus it has no overt subject.[14]

Thus, the Subject Condition may not be universal as a general rule — although it does hold as an overall tendency, since subjectless constructions, if there are any, are usually found only at the periphery of grammar.

4.2.3 Subjects in non-accusative languages

The universality of subjects can also be questioned in a different way: Does the same mapping between thematic roles and GFs obtain in all languages? This has

---

[14]Lowe et al. acknowledge that, if (x)COMP is assumed not to exist as a separate GF (see Section 4.3.1), the clause itself will have to be treated as SUBJ.

long been debated in the literature concerning ergative and other non-accusative types of alignment. Most ergative languages are in fact only morphologically ergative, that is, have ergative case marking while syntactically behaving in the same way as accusative languages. But there are a few languages that have been claimed to be consistently syntactically ergative, e.g. Dyirbal (Dixon 1979), although this analysis is disputed, see Legate (2012); more exotic syntactic alignment types are attested as well. These facts call for an adjustment to the standard approach to argument mapping.

There are two basic proposals for treating non-accusative languages in LFG. One, developed in Manning (1996), is to preserve the standard set of GFs but map SUBJ and OBJ to thematic roles in different languages in different ways. Thus, while intransitive verbs always have a single SUBJ argument, transitive verbs in accusative languages map agents to SUBJ and patients to OBJ (47a); in ergative languages, the mapping is reversed (47b).

(47) a. accusative

    eat ⟨ ag   pt ⟩
           | |
         SUBJ OBJ

  b. ergative

    eat ⟨ ag   pt ⟩
            ╲ ╱
             ╳
            ╱ ╲
         SUBJ OBJ

Thus, in ergative languages, the transitive agent (A) is OBJ while the transitive patient (P) is SUBJ. This explains why the patient has subject-like properties in various constructions. Calling the agent a "direct object" is unfamiliar and confusing from a traditional perspective, which is why Manning proposes an alternative nomenclature of PIVOT (= SUBJ) and CORE (= OBJ, for CORE argument) instead.

This approach works well for languages where one of the arguments fully "takes over" all syntactic properties of subjecthood. However, such languages are an exception rather than the norm. More commonly, subject properties are distributed between the "accusative" subject (A) and the absolutive argument (P): some constructions are aligned in the ergative way, while others are still oriented towards A. For example, in Ashti Dargwa (field data), gender agreement on the verb follows the ergative pattern (S/P), and can even be long-distance (48), which suggests syntactic ergativity. But reflexive binding still prefers the A argment, as in accusative languages (49).

Oleg Belyaev

(48) Ashti (Dargwa > East Caucasian)

di-l   [šin   **d**-eč:-ib]   ha<**d**>eχʷ-i
I(M)-ERG   water(NPL) NPL-drink.PFV-PCVB <NPL>finish.PFV-PRET
'I finished drinking water.'

(49) a. rasul-li **sin-na sa-w**   w-aˤqˤ.aqˤ-ip:i
        R.-ERG self-GEN self-M(ABS) M-hurt.PFV-PRF.3
        'Rasul (erg.) hurt **himself** (abs.).'

    b. * **sin-na   sin-dil** rasul   w-aˤqˤ.aqˤ-ip:i
        self-GEN self-ERG R.(ABS) M-hurt.PFV-PRF.3
        (intended translation: 'Rasul hurt himself.'; lit. 'Himself (erg.) hurt
        Rasul (abs.).')

Falk (2006) observes that cross-linguistically, subject properties tend to fall
into two classes exactly along these lines: anaphoric prominence, switch-reference,
null expression, control of PRO (anaphoric control) and some other properties
such as the ability to serve as the imperative subject are almost always tied to A/
S, even in ergative languages. At the same time, properties related to cross-clausal
continuity – functional control, extraction properties, long-distance agreement
– and certain secondary properties (external structural position in non-configur-
ational languages, agreement) may be tied to different arguments of the clause
in different languages.

Accordingly, Falk proposes to recast the traditional LFG grammatical function
SUBJ as ĜF, which is the most prominent argument (A/S), while introducing the
additional clausal continuity function PIVOT, which can be identified with either
ĜF or OBJ. Subjecthood properties are distributed between these two functions
along the lines in (50).

(50)   Subject properties according to Falk (2006)

ĜF	PIVOT
anaphoric prominence	extraction
anaphoric control	functional control
switch-reference	long-distance agreement
null expression	obligatory element
imperative subject	"external" structural position

Of these two functions, only ĜF can be properly called a grammatical function:
it replaces SUBJ in the argument lists of PRED feature values; in terms of Lexical
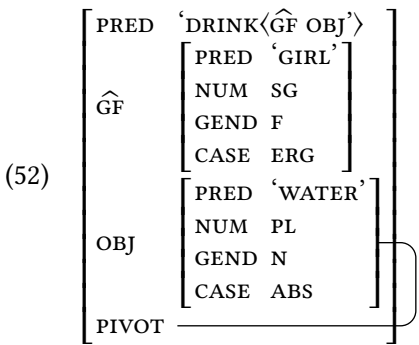Mapping Theory, it is this function that the most prominent argument on the

semantic role hierarchy is mapped to. PIVOT always has to be structure-shared with one of the verbal arguments and is thus more correctly characterized as an overlay function (see Section 5).

All the diverse surface manifestations of PIVOT can be generalized in what Falk calls the Pivot Condition, informally summarized in (51). This condition means that all cross-clausal dependencies, if they are not stated in terms of special overlay functions for long-distance dependencies such as DIS (for "dislocated", or TOPIC and FOCUS in earlier approaches: see Section 5 and Kaplan & Zaenen forthcoming [this volume]) must be tied to PIVOT. Thus PIVOT is the locus through which argument information is shared across clauses.

(51) Pivot Condition:
      A path inward through f-structure into another predicate-argument
      domain or sideways into a coordinate f-structure must terminate in the
      function PIVOT. (Falk 2006: 78)

In English, and in other purely accusative languages, $\widehat{GF}$ and PIVOT are always occupied by the same f-structure. Falk calls such systems "uniform-subject". In other languages, these do not always coincide — this class of languages is called "mixed-subject". The mixed-subject class is not uniform. Its most widespread members are ergative languages, where PIVOT is identified with $\widehat{GF}$ in intransitive clauses and with OBJ in transitive clauses.

Given the facts in (48)–(49), Ashti Dargwa can be analyzed as a mixed-subject, ergative language, with the f-structure of a transitive sentence 'the girl drank water' as in (52).

(52)
$$
\begin{bmatrix}
\text{PRED} & \text{'DRINK}\langle\widehat{GF}\ \text{OBJ'}\rangle \\
\widehat{GF} & \begin{bmatrix} \text{PRED} & \text{'GIRL'} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \\ \text{CASE} & \text{ERG} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'WATER'} \\ \text{NUM} & \text{PL} \\ \text{GEND} & \text{N} \\ \text{CASE} & \text{ABS} \end{bmatrix} \\
\text{PIVOT} & \text{———}
\end{bmatrix}
$$

The Philippine type of alignment, illustrated in (26) above, where any argument can become the "subject" through voice morphology, is interpreted by Falk as promotion to PIVOT, as in (53); the most prominent argument, $\widehat{GF}$, does not change its mapping.

Oleg Belyaev

(53) "Active voice": $(\uparrow \text{PIVOT})=(\uparrow \widehat{\text{GF}})$
       "Direct object voice": $(\uparrow \text{PIVOT})=(\uparrow \text{OBJ})$
       "Indirect object / locative voice": $(\uparrow \text{PIVOT})=(\uparrow \text{OBJ}_\theta)$
       "Instrumental voice": $(\uparrow \text{PIVOT})=(\uparrow \text{OBL}_{\text{INS}})$
       ...

Some languages do not entirely fit the uniform- vs. mixed-subject distinction. In topic prominent languages like Acehnese, PIVOT is identified with any of the core arguments ($\widehat{\text{GF}}$ and OBJ) provided that it bears the information structure function TOPIC. Thus Falk's approach does not require PIVOT to be necessarily tied to particular argument functions.

4.2.4 Universality

Since Falk's framework splits the traditional SUBJ into two grammatical functions that may be identified with different arguments in different languages and constructions, it follows that the subject in the traditional sense – i.e. a single highest-ranking grammatical function that dominates all syntactic rules and processes – is not universal. But we may also ask whether $\widehat{\text{GF}}$ and PIVOT are universal. There are two ways in which a language may be said to lack $\widehat{\text{GF}}$. One is that this language encodes thematic roles directly in the syntax. Such claims have been made for different languages in the literature, especially in the typological tradition. Falk (2006: 169) observes that in LFG terms, this amounts to saying that the language only has oblique GFs: OBL_{AGT}, OBL_{PAT}, etc. This, in turn, entails that the language would have no distinction between core and non-core arguments — a prediction that has empirical consequences. Evaluating such a possibility for Acehnese, one language that has been claimed to lack reference to grammatical relations in its grammar (Van Valin & LaPolla 1997), Falk concludes that its syntax does distinguish core functions from non-core functions and thus requires reference to $\widehat{\text{GF}}$. Similarly, Kibrik (2000), as mentioned in Section 3.4.3, argues that most constructions in Archi (Lezgic > East Caucasian) are only sensitive to the term (core argument) vs. non-term distinction. But there is one construction in Archi that is oriented towards A/S arguments (i.e., in Falk's terms, $\widehat{\text{GF}}$): clausemate reflexivization. Nevertheless, the theoretical possibility of languages only having oblique arguments still exists and deserves to be investigated in more detail, although, based on the current state of our understanding, their existence does not seem likely.

Another sense in which a language may lack $\widehat{\text{GF}}$ is, conversely, if it draws a more fine-grained distinction between core arguments, i.e. does not unify the

transitive agent (A) and the sole intransitive argument (S) in any way. Again, this approach is widespread in the typological / functionalist tradition, a prominent example being Dixon (1994), who treats A, S and P as syntactic primitives. This is useful for purposes of typology and cross-linguistic comparison: A, S and P serve as valid comparative concepts in the sense of Haspelmath (2009). But applied to individual grammars, this distinction seems too fine-grained, failing to capture important generalizations. It is well-known, for example, that $\widehat{\text{GF}}$ outranks other arguments in anaphoric constructions in the overwhelming majority of languages, regardless of their other alignment patterns. Nor do "syntactically tripartite" languages with S, A and P having distinct, non-intersecting sets of properties seem to be attested.

Thus, $\widehat{\text{GF}}$ is likely to be universal. A separate question is what a PIVOTLESS language could look like, and whether such languages exist. A pivotless language is *not* a language where the pivot cannot be readily identified with any grammatical function; it could be identified with the topic, as in Acehnese, or with the highest-ranking argument on the person hierarchy, as in some analyses of Ojibwe (Algonquian > Algic, Rhodes 1994). A pivotless language would rather lack constructions of the kind that are predicted to be pivot-sensitive by the Pivot Condition (51). For example, there would be no cross-clausal extraction, with all interrogatives and relatives being localized in their local domains; coreference in coordination and in other multiclausal constructions would similarly involve no pivot sensitivity. Falk argues that at least two languages, Choctaw and Warlpiri, qualify for pivotless status.

Falk's approach is insightful and makes a number of strong claims that deserve more thorough cross-linguistic investigation. It is widely accepted as the most adequate solution for ergativity and other syntactic alignments within LFG, although many authors still continue using the SUBJ-OBJ distinction for languages where Falk's fine-grained analysis is irrelevant, i.e. mainly in syntactically accusative languages. Falk's notion of $\widehat{\text{GF}}$ and PIVOT also has yet to be fully integrated with the recent developments in the relevant areas of LFG, such as Lexical Mapping Theory and semantic composition.

4.3 Complementation and nonverbal predication

In the preceding sections, I have mostly avoided discussing sentential complements, because their specialized grammatical function COMP stands apart from other grammatical functions in LFG. It is not readily classifiable in terms of termhood and semantic restrictedness, and its limitation to a single semantic type (clauses / states of affairs) is unusual for LFG. In fact, the very existence of

Oleg Belyaev

(x)COMP as a separate GF has been questioned in the theoretical literature, as discussed in Section 4.3.1. In Section 4.3.2, I discuss the difference between closed (COMP) and open (XCOMP) complements. Nonverbal predication is also sometimes analysed using the grammatical function XCOMP, and therefore it is discussed under the same umbrella in Section 4.3.3.

4.3.1 The status of COMP

The status of COMP as a specialized grammatical function in LFG is controversial. From the beginning, it was assumed that *all* clausal complements are classified as COMP (Kaplan & Zaenen 1989, Bresnan et al. 2016). As a formal assumption, this idea is suspect: the spirit of LFG is generally to separate categorial and functional information, such that f-structure should not draw a distinction between NP and CP complements. For this reason, the very existence of COMP has been questioned, first in Alsina et al. (1996), who proposed that COMP can be replaced by OBJ.

One argument in favour of COMP is the fact that it can coexist with OBJ and OBJ$_\theta$, as in (54).

(54) David bet [Chris]$_{\text{OBJ}}$ [five dollars]$_{\text{OBJ}_{\text{GOAL}}}$ [that he would win]$_{\text{COMP}}$.

As a further argument, Dalrymple & Lødrup (2000) show that while many clausal complements in English, German and Swedish do, indeed, behave like objects, others do not. For example, in German the complement of the verb 'believe' can be replaced by a pronoun and moved to clause-initial position (55); the latter option is also available for ordinary object NPs (56). In contrast, neither option is possible for complements of 'be happy' (57).

(55) German (Germanic > Indo-European)

    a. Ich glaube [**dass** die Erde rund ist].
       I   believe   that the earth round is
       'I believe **that** the earth is round.'

    b. Ich glaube es.
       'I believe **it**.'

    c. [**Dass** Hans krank ist] glaube ich.
        that Hans sick   is believe I
       '**That** Hans is sick, I believe.'

(56) Einen Hund habe ich gesehen.
      a   dog   have I   seen
      'A dog, I have seen.'

(57) a. Ich freue   mich   [**dass** Hans krank ist]
          I   gladden myself that Hans sick   is
          'I am happy that Hans is sick.'

      b. * Ich freue mich **das** / **es**.
           ('I am happy it.')

      c. * [**Dass** Hans krank ist] freue ich mich.
           ('**That** Hans is sick, I am happy.')

Dalrymple & Lødrup (2000) conclude that while clausal arguments of verbs like 'believe' do indeed bear the function OBJ in German, complements of verbs like 'be happy' should be recognized as genuine COMPs. However, Alsina et al. (2005) contest this conclusion by appealing to the data of Catalan and Spanish. They claim that both examples like (54) and the data cited by Dalrymple and Lødrup only show that OBJ alone is not enough to capture the behaviour of all types of clausal complements. But if some complements are treated as OBJ$_\theta$ and OBL$_\theta$, they can coexist with direct objects, and their syntactic properties can be adequately captured. A similar conclusion is reached in Forst (2006) for the German data.

This debate still continues in the LFG literature. Thus Belyaev et al. (2017) conclude that the syntax of complementation in Moksha Mordvin requires appealing to COMP in addition to OBJ and OBL$_\theta$. Moksha has object agreement morphology on transitive verbs. As discussed in Section 3.4 above, agreement is a feature of terms; clausal complements controlling agreement may thus be viewed as OBJ. In Moksha, there is a split according to this criterion. Factives control agreement, and they can also be replaced by pro-forms (58), like OBJ-complements in German, passivized, coordinated with nouns, and replaced by quantificational expressions.

(58) Moksha Mordvin (Mordvinic > Uralic)
      factive complements

      a. object agreement
         učit'əl'-s'   **soda-si-n'ə** /   *soda-s'   [**što**
         teacher-DEF.SG[NOM] know-NPST.3PL.O.3SG.S know-NPST.3SG COMP
         pet'ɛ er'   mejn'ɛ   vor'g-əčn'-i   urok-stə]
         Peter every what.TMPR run.away-IPFV-NPST.3SG class-EL
         'The teacher **knows** (SUBJ-OBJ) **that** Peter always misses classes.'

Oleg Belyaev

    b. pronominalization

       mon   kunarə   soda-jn'ə   [što   vas'ɛ
       I[NOM] for.a.long.time know-PST.3.O.1SG.S   COMP Basil[NOM]

       ašč-əl'   t'ur'ma-sə] — də mon-gə **t'ɛ-n'**   soda-sa
       be-PQP.3SG prison-IN   yes I-ADD this-GEN know-NPST.3SG.O.1SG.S

       'I have known (SUBJ-OBJ) for a long time that Basil had been in prison.
       – Yes, I know (SUBJ-OBJ) **it** too.'

Other complement clauses do not control matrix verb agreement, i.e. the verb only agrees with the subject. However, this class is not homogeneous. Some non-factive complements, such as the complement of 'fear', can be replaced by pronominal postpositional phrases or oblique case-marked pronouns — these can uncontroversially be classified as obliques (59). But complements of other non-factives, such as the verb 'say', cannot be replaced by a pronoun — an adverbial 'so' should be used instead (60). They also cannot be replaced by quantificational expressions or coordinated with a nominal argument. Belyaev et al. (2017) conclude that this latter class of complements, being distinct from both OBJ and OBL$_\theta$, should be assigned the grammatical function COMP.[15]

(59)   Moksha Mordvin (Mordvinic > Uralic)
      non-factive 'fear': pronominalization
      mon   dumand-an   [što   vas'ɛ   af pastupanda-v-i]
      I[NOM] think-NPST.1SG COMP Basil[NOM] NEG enter-PASS-NPST.3SG

      institut-u — mon   tožə **t'a-də** pel'-an
      institute-LAT I[NOM] also that-ABL fear-NPST.1SG

      'I think (SUBJ) that Basil will not enter the university. — I am afraid (SUBJ) **of that** as well.'

(60)   non-factive 'say': no pronominalization

---

[15]Another option is available: these non-agreeing complements can be OBJ$_\theta$. This idea is appealing because Dalrymple & Nikolaeva (2011) analyze unmarked direct objects in DOM systems as OBJ$_\theta$. In Moksha, which displays DOM, direct objects can be nominative (unmarked) or genitive. Indeed, it is unmarked direct objects in Moksha that are similar to complements of verbs like 'say': they do not trigger agreement, cannot be used with quantifiers; pronominal objects are always case-marked, etc. However, it is not clear whether unmarked and genitive direct objects in Moksha should be assigned to different grammatical functions: for instance, a marked and unmarked direct object can be coordinated (Natalia Serdobolskaya, p.c.). In contrast, complements of verbs like 'say' cannot be coordinated with a noun phrase (Belyaev et al. 2017). Thus for Moksha the answer depends on whether unmarked direct objects in this language are OBJ$_\theta$s and on whether the coordination facts can be given an alternative explanation.

40

nu   mon   **t'aftə** / **t'ɛ-n'**   af   dumand-an
well I[ɴᴏᴍ] thus   this-ɢᴇɴ ɴᴇɢ think-ɴᴘꜱᴛ.1ꜱɢ
{Context: 'Basil is so smart, he will surely pass the exams with excellent
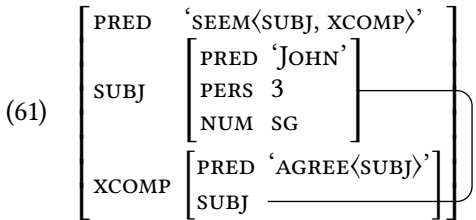marks! —} Well, I do not think (ꜱᴜʙᴊ) **so** / *that.'

Not all languages with object agreement or indexing draw such a sharp distinction between different complement types, however. West Circassian (West Caucasian, polysynthetic), for example, treats most clausal complements in the same way as NP arguments, which is consistent with this language's weak distinction between nouns and verbs (Letuchiy 2016).

Significant differences between clausal complements and "nominal" grammatical functions such as ꜱᴜʙᴊ and ᴏʙᴊ have also been described for Russian in Letuchiy (2012). Overall, the data strongly suggest that ᴄᴏᴍᴘ should at least be recognized as a possible ɢꜰ for clausal complements, although the extent to which languages use this possibility seems to vary. The semantic differences between ᴏʙᴊ and ᴄᴏᴍᴘ complement clauses should also be investigated in more detail.

4.3.2 Open and closed complements

We mentioned above that clausal complements in LFG are split into two grammatical functions: ᴄᴏᴍᴘ and xᴄᴏᴍᴘ. The former is called "closed", the latter "open". This means that the former are, in principle, fully self-contained and have their own subjects (e.g. finite complement clauses); the latter do not have a subject, which has to be structure shared with an argument of another clause. Open complements (xᴄᴏᴍᴘ) appear in structures called ꜰᴜɴᴄᴛɪᴏɴᴀʟ ᴄᴏɴᴛʀᴏʟ, which involves structure sharing of an argument of the matrix clause and an argument (usually the subject) of the subordinate clause. Functional control is generally used to represent so-called raising constructions, as in (38), repeated here, with the f-structure in (61).

(38)   **John** seemed [__ to agree].

(61)
$$
\begin{bmatrix}
\text{PRED} & \text{'SEEM}\langle\text{SUBJ, XCOMP}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'JOHN'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'AGREE}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \text{\underline{\hspace{2cm}}} \end{bmatrix}
\end{bmatrix}
$$

Oleg Belyaev

Functional control in LFG is opposed to anaphoric control, which is often employed to analyse the construction known as Equi or simply control in English, see (62).[16]

(62)   Chris told John$_i$ [(PRO$_i$) to come tomorrow].

Anaphoric control involves no structure sharing but only a covert pronominal subject in the subordinate clause (PRO); accordingly, complements whose subject is anaphorically controlled are treated as closed (COMP). The f-structure of (62) is shown in (63), where the dashed line indicates coreference.

(63)
$$
\begin{bmatrix}
\text{PRED} & \text{‘TELL}\langle\text{SUBJ, OBJ, COMP}\rangle\text{’} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘CHRIS’} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘JOHN’} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\
\text{COMP} & \begin{bmatrix} \text{PRED} & \text{‘COME}\langle\text{SUBJ}\rangle\text{’} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘PRO’} \end{bmatrix} \\ \text{ADJ} & \{[\text{PRED} \ \text{‘TOMORROW’}]\} \end{bmatrix}
\end{bmatrix}
$$
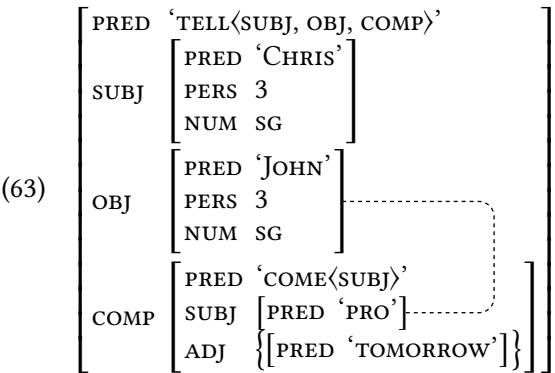
It is not clear if the distinction between COMP and XCOMP is really needed to account for the behaviour of control constructions. After all, equations that enforce structure sharing automatically ensure that the subject of the complement clause is overtly expressed only once: double expression would cause a PRED conflict. F-structure does not take the linear order or c-structure position of elements into account, therefore it does not matter, in principle, *where* the argument is expressed. This means that LFG allows Backward Raising constructions as in the West Circassian (64) by default (Sells 2006). In (64), the "raised" NP is overtly expressed only in the subordinate clause, which is seen in its case marking: the ergative is selected by the verb 'lead'. The main clause subject, if it were overt, would have been in the absolutive (as seen in the crossed out pronoun).

(64)   West Circassian (West Caucasian)

---

[16]The discussion here presents a simplified view of the issue. In some LFG work, functional control is not limited to raising constructions but is also used in the analysis of some or all of the constructions traditionally called Equi or control. See Vincent forthcoming [this volume] for detailed information on control and raising in LFG.

a-xe-r̶ [a-xe-me se s-a-š'e-new]
DEM-PL-ABS DEM-PL-ERG.PL 1SG.ABS 1SG.ABS-3PL.ERG-lead-INF
∅-fjež'a-ʁe-x
3ABS-begin-PST-3PL.ABS
'They began to lead me.' (Potsdam & Polinsky 2012: 76)

The English counterpart to this example would be *Began [*they to lead me*] (or, to provide an uncontroversial example of raising, *Seem they to come*). The ungrammaticality of such examples requires independent explanation (for example, English xcomps are expressed by VPs at c-structure, which do not have a subject position). See Vincent forthcoming: §7 [this volume] for further discussion of LFG analyses of backwards raising.

Similarly, anaphoric control is typically analyzed as coreference that is syntactically enforced through equations like (↑ SUBJ INDEX) = (↑ COMP SUBJ INDEX)[17] and, possibly, (↑ COMP SUBJ PRED)='PRO'. If the latter equation is present, an overt subject in the complement clause is precluded due to PRED conflict. If it is not, argument expression is only constrained by general anaphoric requirements, which is why Backward Control (Polinsky & Potsdam 2002) is impossible in most languages due to Principle C violations (see Rákosi forthcoming [this volume] for details on Principle C).

Crucially, such constraints follow from universal considerations, functional equations and general principles of individual grammars, but not from complements being xcomp rather than comp. Thus, it is not clear whether the traditional distinction between comp and xcomp is anything more than a useful notational convention; both could be said to refer to the same GF.

4.3.3 Nonverbal predication

Traditionally, xcomp was used in LFG to represent nonverbal predicates, treating them as arguments of copular verbs such as 'to be', as in (65).

(65)   a. John is kind.

---

[17]In an approach where coreference is a semantic relation, such as Haug (2013), it cannot be enforced directly in the f-structure, but it can be done via a Glue meaning constructor (Haug 2014, Asudeh forthcoming [this volume]).

Oleg Belyaev

b.
$$\begin{bmatrix} \text{PRED} & \text{`BE}\langle\text{SUBJ, XCOMP}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`JOHN'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`KIND}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \end{bmatrix} \end{bmatrix}$$
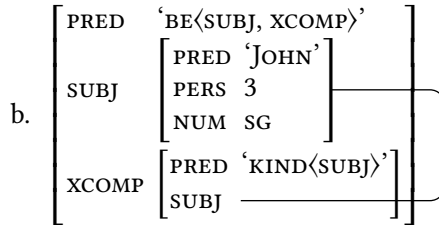
This effectively makes the nonverbal predicate into a kind of small clause. The main problem with this approach is that all lexical items that can serve as predicates must have two subcategorization frames, because in normal contexts at least nouns, and possibly adjectives (if they are not assumed to be predicated of their head noun), do not have a valency for SUBJ. As observed in Dalrymple et al. (2004), another problem for this approach is that clauses that already have subjects may function as predicates, as in the sentence *The problem is that John came.* Such clauses have no open subject position to share with the matrix subject.

The main alternative is to replace XCOMP with a special grammatical function PREDLINK (Butt et al. 1999), which is not an open complement GF and therefore does not have to share a subject valency, see (66) for *John is kind* and (67) for *The problem is that John came.*

(66)
$$\begin{bmatrix} \text{PRED} & \text{`BE}\langle\text{SUBJ, PREDLINK}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`JOHN'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{`KIND'} \end{bmatrix} \end{bmatrix}$$

(67)
$$\begin{bmatrix} \text{PRED} & \text{`BE}\langle\text{SUBJ, PREDLINK}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`PROBLEM'} \\ \text{DET} & \text{DEF} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{`COME}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`JOHN'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$
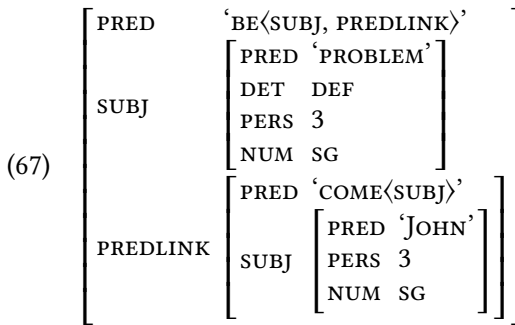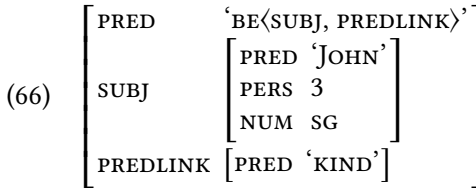
One drawback of the PREDLINK approach compared to the XCOMP approach is related to the fact that in languages with adjective agreement, such as Russian (68), the predicative adjectives agree in gender with the subject. In (68a), the word *komnata* 'room' is feminine, and therefore the predicative adjective *malen'kaja* is feminine. In (68b) *dom* 'house' is masculine, and the adjective is also masculine.

(68) Russian (Slavic > Indo-European)

   a. Komnata   byla   malen'kaja.
     room(F).SG.NOM was.F.SG small.F.SG.NOM
     'The room was small (f.).'

   b. Dom   byl   malen'kij.
     house(M).SG.NOM was.M.SG small.M.SG.NOM
     'The house was small (m.).'

This is straightforward to capture in the XCOMP approach, because the adjective has its own local subject with which it can agree: see (69).

(69)
$$
\begin{bmatrix}
\text{PRED} & \text{'BE}\langle\text{SUBJ, XCOMP}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'ROOM'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'SMALL}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \underline{\quad\quad} \end{bmatrix}
\end{bmatrix}
$$
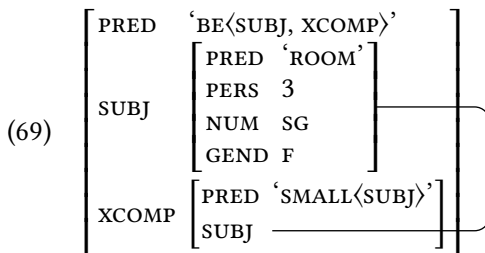
Adnominal adjectives like in (70a) can be treated in the same way by using a cyclic f-structure (70b) (see Haug & Nikitina 2012), requiring only one agreement pattern in the lexical entry (71).

(70) Russian (Slavic > Indo-European)

   a. malen'kaja   komnata
     small(F).SG.NOM room(F).SG.NOM
     'small room'

   b.
$$
\begin{bmatrix}
\text{PRED} & \text{'ROOM'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{SG} \\
\text{GEND} & \text{F} \\
\text{XADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'SMALL}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \underline{\quad\quad} \end{bmatrix} \right\}
\end{bmatrix}
$$

(71)   *malen'kaja* A   ($\uparrow$PRED) = 'SMALL$\langle$SUBJ$\rangle$'
                    ($\uparrow$ SUBJ NUM) = SG
                    ($\uparrow$ SUBJ GEND) = F

    In the PREDLINK approach, agreement rules will have to be more complex, utilizing inside-out functional expressions as in (72a) for adnominal adjectives and (72b) for predicative adjectives.

Oleg Belyaev

(72) a. ((ADJ ∈ ↑) NUM) = SG

  b. ((PREDLINK ↑) SUBJ NUM) = SG

Yet another approach is to unify the f-structure of the nonverbal predicate with the f-structure of the clause (via ↑=↓); this is proposed in Dalrymple et al. (2004) for languages like Japanese, where predicative adjectives do not require a copula (73).

(73) Japanese (Japonic)

  a. hon wa   akai
    book TOPIC red
    'The book is red.'

  b. $\begin{bmatrix} \text{PRED } \text{'RED}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ } \begin{bmatrix} \text{PRED } \text{'BOOK'} \end{bmatrix} \end{bmatrix}$

In Japanese, this analysis is quite reasonable because adjectives are morphologically a subclass of verbs. It is plausible to assume that even adnominal adjectives have subjects, and thus always have PRED values like 'RED⟨SUBJ⟩'. But for languages like Russian, where adjectives inflect like nouns, there is less evidence in favour of treating each adjective as having a subject. Therefore, this analysis suffers from the same disadvantage as the XCOMP approach, in requiring two lexical definitions for each adjective or noun. Apart from this, it is structurally quite distinct from both the XCOMP and the PREDLINK approaches in being monostratal. Overall, as Dalrymple et al. (2004) conclude, it is likely that all three approaches are required to account for different constructions in different languages. For more information on copular constructions in LFG, see Dalrymple et al. (2019: 189–197).

4.3.4 The classification of COMP

4.3.4.1 Termhood   The termhood of sentential complements has not been frequently discussed in the literature. In no small part this is due to the unclear status of the grammatical function COMP itself (see Section 4.3.1 above). A number of arguments in favour of treating COMP as a non-term GF are given in Dalrymple et al. (2019). If this view is combined with the idea that clausal complements are split between COMP and OBJ (Dalrymple & Lødrup 2000), one can predict that in languages with object agreement, OBJ-like complements may trigger agreement on the verb while COMPs may not. This prediction is confirmed in languages like Moksha Mordvin (Mordvinic > Uralic), where, as Belyaev et al. (2017) argue,

the verb agrees with OBJ-like complements (mainly those of factive verbs like 'know') but does not agree with COMP-like complements (mainly propositional complements of verbs like 'promise'):

(74) Moksha Mordvin (Mordvinic > Uralic)
   učit'əl'-s'   **soda-si-n'ə /**   *soda-s'   [**što**
   teacher-DEF.SG[NOM] know-NPST.3PL.O.3SG.S know-NPST.3SG COMP
   pet'ɛ   er'   mejn'ɛ   vor'g-əčn'-i   urok-stə]
   Peter every what.TMPR run.away-IPFV-NPST.3SG class-EL
   'The teacher **knows** (SUBJ +OBJ) **that** Peter always misses classes.'

(75) paša   **abəščanda-s'** /   *abəščanda-z'ə   [**što**
   Paul[NOM] promise-PST.3SG promise-PST.3SG.O.3SG.S COMP
   il'caman'   kud-u]
   accompany.NPST.1SG.O.3SG.S house-LAT
   'Paul **promised** (SUBJ) **that** he would accompany me home.'

4.3.4.2 Semantic restrictedness   The status of (x)COMP as semantically restricted is less clear. Certainly, sentential complements are semantically diverse: at least factives and non-factives have been distinguished since Kiparsky & Kiparsky (1970), and other distinctions since then have been discussed in the literature, such as between fact, proposition, event (Peterson 1997) and other abstract objects (Asher 1993). However, this is a difference in the semantic type of the argument and its entailments/presuppositions, which is not directly related to semantic roles; it might be more properly compared to the distinction between definite and indefinite NPs — given that definites, like factives, presuppose the existence of their referents, and have other similar properties (see Melvold 1991).

The range of semantic roles that clausal arguments can be associated with is difficult to resolve because these arguments are rather restricted in their distribution. There are very few verbs with two clausal arguments (exceptions being verbs like *prove*, *entail*, etc.), and these all have only SUBJ and COMP arguments; I am not aware of any verbs that have two sentential non-subjects (COMP, OBJ or OBL). Clausal arguments often cannot have the markings characteristic of NP arguments and hardly ever undergo valency-changing processes (even clausal complements classified as OBJ can be difficult to passivize). Hence, there is little distributional evidence that could help distinguish between the semantic roles of COMP. On a purely speculative basis, one may say that most COMPs are Themes, some are Stimuli (mental predicates), and some could be classified as Goals (e.g. verbs like *try*). In terms of Dowty (1991), these all fall under the proto-role Patient;

thus it is an open question whether these fine-grained distinctions are grammatically relevant. Zaenen & Engdahl (1994) believe that they are not, and that (x)comp is, in fact, semantically restricted, since this GF can only be occupied by clausal arguments. Similarly, Dalrymple & Lødrup (2000), who distinguish between comp and obj (see Section 4.3.1), assume that comp is semantically restricted and that this is what distinguishes comp from obj.

The alternative is simply to avoid definitively classifying comp and xcomp as either semantically restricted or semantically unrestricted. Falk (2001) proposes that comp and xcomp are different from all other GFs in having the positive value for a special feature [c] (for complement). In practical terms, this is equivalent to the position of Zaenen & Engdahl (1994). Another approach is to treat comp as underspecified for being semantically restricted or unrestricted, depending on the context, as in Berman's (2007) analysis of German.

The difficulties in resolving this question only serve to illustrate that comp and xcomp are really apart from all other GFs and require a special analysis – if they are to be distinguished at all, as discussed in detail in Section 4.3.1 above.

4.4 Possessors

The discussion of grammatical functions has so far avoided mentioning possessors. This is because, being nominal dependents, they are not easily comparable to other, clause-level GFs.

In LFG, possessors are standardly assumed to bear the grammatical function poss. Among clausal GFs, it is most similar to subj in two ways. First, it is the most prominent argument, as, apart from possessors, nouns may only have oblique dependents. Second, it is semantically unrestricted. It is well-known that possessors (in the syntactic sense, i.e. genitive dependents) can have a very wide range of relations to their heads. The semantic non-restrictiveness of possessors is also evident from the fact that in many languages, genitive marks the same arguments in non-finite clauses that are mapped to subj in finite clauses (76).

(76)   a. **The enemy** destroyed the city.
       b. **the enemy**'s destruction of the city

Therefore, some authors propose reducing poss to subj (Sulger 2015). This solution seems too radical, however — at least for some languages. Chisarik & Payne (2003) were the first to introduce a hybrid approach that uses both subj and poss in noun phrases. They analyse English and Hungarian, which both allow two types of possessor expression: English has the "Saxon Genitive" *'s* and *of*-possessors, while Hungarian has nominative and dative possessors. Chisarik and

Payne argue that English *'s*-possessors and Hungarian nominative possessors are SUBJS, while the other two types of possessors are ADNOMS, which correspond to POSS. Laczkó (2004), critical of their analysis of the Hungarian data, also maintains that Hungarian possessors can be either SUBJ or POSS, but argues that the GF of the possessor is independent of its marking pattern. Laczkó further develops this analysis of Hungarian in a series of papers, in particular Laczkó (2009, 2017). Laczkó & Rákosi (2019) further argue that in some Hungarian examples such as (77), both SUBJ and POSS are present in the f-structure of the nominalization. In this case, the possessor is the reciprocal which triggers 3rd person singular agreement on the nominalized verb, while the subject is the null pronominal coreferent with 'boys' in the main clause (78).

(77) A fiúk$_i$ dijazzák   [$_{DP}$ az egymás$_i$ lefest-és-é-t].
      the boys appreciate.3PL   the each.other paint-NMLZ-POSS.3SG-ACC
      'The boys appreciate the painting of each other.' (Laczkó & Rákosi 2019:
      163)

(78)
$$\begin{bmatrix} \text{PRED} & \text{'APPRECIATE}\langle\text{SUBJ OBJ}\rangle\text{'} \\ \text{SUBJ} & [\text{"the boys"}] \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'PAINTING}\langle\text{SUBJ POSS}\rangle\text{'} \\ \text{SUBJ} & [\text{"pro"}] \\ \text{POSS} & [\text{"each other"}] \end{bmatrix} \end{bmatrix}$$
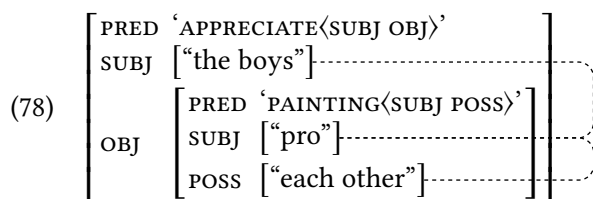
If POSS is a governable GF like SUBJ, all nouns with optional possessors must be assumed to have two variant PRED values: with and without a possessor valency, e.g. 'book' and 'book-of⟨POSS⟩' (Bresnan 2001, Bresnan et al. 2016: 315 et passim). This seems undesirable, so Dalrymple et al. (2019) propose to treat POSS as being ungovernable, like ADJ, but positioned at the top of the GF hierarchy, like SUBJ. This means that POSS is licensed in any f-structure having a PRED value, including clausal f-structures; thus, additional care must be taken to ensure that POSS is constrained not to appear in inappropriate positions.

5 Overlay and discourse functions

F-structures occupying GF feature values may have additional functions in the clause that link the f-structure to the wider syntactic or discourse context. Following Falk (2001: 59), who took the term from Johnson & Postal (1981), these can be called OVERLAY FUNCTIONS because they must always be connected to arguments or adjuncts by either anaphora or structure sharing (according to

Oleg Belyaev

Extended Coherence, see Fassi Fehri 1988, Zaenen 1985, Bresnan & Mchombo 1987, Bresnan et al. 2016: 62–63). One overlay function, PIVOT, serves to capture some of the subject properties of core arguments and has been discussed in Section 4.2.3. Two other important classes of functions are so-called grammaticalized discourse functions, which traditionally included TOPIC and FOCUS but are now increasingly replaced by a single function called DIS or UDF (discussed in Section 5.1), and functions like Q or RELPRO that are intended to mark elements relativized, questioned, or otherwise selected to serve as input to other syntactic or semantic processes (discussed in Section 5.2).

5.1 TOPIC, FOCUS and DIS

Since the earliest work in LFG, "grammatic(al)ized discourse functions" TOPIC and FOCUS have been used at f-structure to represent simultaneously the information structure status of participants and their role in establishing long-distance dependencies such as wh-extraction. It is also often assumed, e.g. in Bresnan (2001) and Bresnan et al. (2016), that SUBJ is unique in being both a grammatical function and a discourse function. This is meant to represent the discourse prominence of subjects and capture some generalizations in the c- to f-structure mapping, but it also means that discourse functions in this understanding are not necessarily overlay functions.[18]

Under this view, f-structure combines morphosyntactic and information-structure features, which is against LFG's tendency for localizing different aspects of language structure at different projections or levels (see Belyaev forthcoming(c) [this volume] and Belyaev forthcoming(a) [this volume]). This, with other formal and empirical considerations, has caused recent work, notably King & Zaenen (2004) and Dalrymple & Nikolaeva (2011), to promote information structure to a separate projection (see Zaenen forthcoming [this volume]), which has removed the need to represent notions such as topic and focus at f-structure. Therefore, many authors feel that only one overlay function is now sufficient for all topicalized, focalized or otherwise displaced material. This function has been variously called UDF for "unbounded dependency function" (Asudeh 2012), OP for "operator" (Alsina 2008), or DIS for "dislocated" (Dalrymple et al. 2019) in the

---

[18] Falk (2006), whose approach was discussed in Section 4.2.3 above, introduces the overlay function PIVOT to account for those subject properties that are associated with syntactic prominence. Therefore, the properties that Bresnan et al. associate with SUBJ as a discourse function can instead be associated with PIVOT in Falk's approach, resolving the ambiguous status of subjects. I am grateful for this observation to an anonymous reviewer.

literature.[19]

Regardless of whether DIS or TOPIC / FOCUS are used, these attributes have to be set-valued because there may be multiple dislocated elements in one sentence, whether in the same position, like in (79) from French, where two phrases are right-dislocated (with clitic resumption), or in different positions, as in (80) from English, where *Mary* and *me* are dislocated to the left and right edges of the clause, respectively.

(79)   a.   French (Romance > Indo-European)

Je   le   lui   ai   donné, **le livre à Jean**.
I.CL it.CL to.him.CL have given   the book to J.

b.

$$
\begin{bmatrix}
\text{PRED} & \text{`GIVE}\langle\text{SUBJ, OBJ, OBJ}_{\text{RECIP}}\rangle\text{'} \\
\text{TENSE} & \text{PAST} \\
\text{DIS} & \left\{
\begin{bmatrix}
\text{PRED} & \text{`BOOK'} \\
\text{DEF} & + \\
\text{PERS} & 3 \\
\text{NUM} & \text{SG} \\
\text{GEND} & \text{M}
\end{bmatrix}
\begin{bmatrix}
\text{PRED} & \text{`JEAN'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{SG} \\
\text{GEND} & \text{M} \\
\text{PCASE} & \text{A}
\end{bmatrix}
\right\} \\
\text{SUBJ} & \begin{bmatrix}
\text{PRED} & \text{`PRO'} \\
\text{PERS} & 1 \\
\text{NUM} & \text{SG}
\end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}
\text{PRED} & \text{`PRO'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{SG} \\
\text{GEND} & \text{M} \\
\text{CASE} & \text{ACC}
\end{bmatrix} \\
\text{OBJ}_{\text{RECIP}} & \begin{bmatrix}
\text{PRED} & \text{`PRO'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{SG} \\
\text{GEND} & \text{M} \\
\text{CASE} & \text{DAT}
\end{bmatrix}
\end{bmatrix}
$$

(80)   a.   ***Mary, I saw her yesterday, me.***

---

[19]The treatment of long-distance dependencies in LFG is described in detail in Kaplan & Zaenen forthcoming [this volume]; here, I will only discuss issues related to the role overlay functions play in their analysis.

Oleg Belyaev

$$
\text{b.}\begin{bmatrix}
\text{PRED} & \text{`SEE}\langle\text{SUBJ,OBJ}\rangle\text{'} \\
\text{TENSE} & \text{PAST} \\
\text{DIS} & \left\{\begin{bmatrix}\text{PRED} & \text{`MARY'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG}\end{bmatrix} \begin{bmatrix}\text{PRED} & \text{`PRO'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG} \\ \text{CASE} & \text{ACC}\end{bmatrix}\right\} \\
\text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`PRO'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG} \\ \text{CASE} & \text{NOM}\end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}\text{PRED} & \text{`PRO'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{CASE} & \text{ACC}\end{bmatrix} \\
\text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{`YESTERDAY'}\end{bmatrix}\right\}
\end{bmatrix}
$$

Notice that the f-structures do not distinguish between two types of dislocation: in the DIS approach, all dislocated elements are members of the same set, while in the TOPIC / FOCUS approach, both would be TOPICs due to their information structure status. Presumably, a distinction at f-structure is not required because the difference between types of dislocation is captured at other levels, such as information structure (i-structure) or prosody (p-structure).

In fact, when so much has been delegated to other levels, it is not clear whether it is really necessary to indicate the dislocated status of a constituent by any f-structure feature. Indeed, in all the analyses of long-distance dependencies that I am aware of, DIS is locally introduced in the rule that defines the dislocated position by the equation $\downarrow \in (\uparrow \text{DIS})$, and no other rules reference the value of DIS directly. The symbol GF used in paths constraining long-distance dependencies usually includes only non-overlay GFs (Dalrymple et al. 2019: 206), so the dislocation of a phrase from one clause to another does not influence its availability for further extraction. When the dislocated phrase *is* relevant for other processes, such as in relativisation and constituent questions, it occupies the special overlay functions RELPRO and Q. It thus appears that the feature DIS duplicates the information already present at c-structure – that the element is in some dislocated position – and is therefore redundant. This question is discussed in detail in Snijders (2015: section 4.6).

5.2 RELPRO and Q

In some constructions, elements that are dislocated to designated structural positions serve as input to other syntactic or semantic rules and constraints. For example, in relative clauses, the relative pronoun must be linked to the head of the relative phrase, both in syntax (e.g. to ensure agreement in gender and/or number) and in semantics (in order to correctly restrict the reference of the head noun). Similarly, the semantic interpretation of constituent questions must be able identify the f-structure of the interrogative.

It is not enough to use only DIS in such constructions because DIS is not specific enough. A sentence may have another dislocated element in addition to the relative pronoun or interrogative: for example, in the sentence *John, who saw him?* the f-structures of both *John* and *who* will be elements of DIS, but only *who* must be correctly identified as the question word. The traditional distinction between TOPIC and FOCUS will not help either, because relativization or questioning of a phrase often leads to the extraction of a larger constituent in which it occurs (pied piping), as in the sentence *Whose brother did John see?*, where the dislocated element occupying FOCUS is *whose brother*, but only *whose* is the interrogative element.

For these reasons, LFG analyses of relativisation and constituent questions make use of the additional overlay features RELPRO and Q, respectively, that specifically include the f-structure of the element that is relativized or questioned.[20] For example, the sentence *Whose brother did John see?* will have the f-structure in (81).

---

[20]Similar effects could be achieved by using off-path constraints (see Belyaev forthcoming(a) [this volume] on the notion) but this seems to be in essence equivalent to using the overlay functions but results in a more cumbersome analysis (Tracy Holloway King, p. c.). This possibility is explored in Kaplan & Zaenen forthcoming: §4 [this volume].

Oleg Belyaev

(81)

$$\begin{bmatrix} \text{PRED} & \text{‘SEE}\langle\text{SUBJ,OBJ}\rangle\text{’} \\ \text{TENSE} & \text{PAST} \\ \text{DIS} & \left\{ \begin{bmatrix} \text{PRED} & \text{‘BROTHER’} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{POSS} & \begin{bmatrix} \text{PRED} & \text{‘PRO’} \\ \text{PRONTYPE} & \text{WH} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{CASE} & \text{GEN} \end{bmatrix} \end{bmatrix} \right\} \\ \text{Q} & \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘JOHN’} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{OBJ} & \end{bmatrix}$$
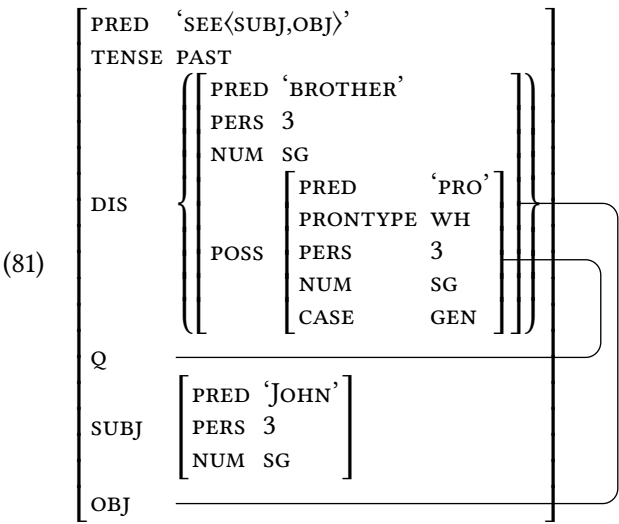
In this example, the question word is the possessor *whose*, but English does not allow extraction of just the possessor, so the whole object phrase *whose brother* is dislocated to the left periphery and, consequently, appears in DIS at f-structure. The wh-word itself occupies the value of the special overlay function Q, which represents the element being questioned. For more information on the handling of long-distance dependencies in LFG, see Kaplan & Zaenen forthcoming [this volume].

6 Conclusions

In this chapter, I have described the key properties of the LFG view of grammatical functions. LFG is unique among contemporary syntactic frameworks in assigning to grammatical functions a central role in grammar, without reducing them to more basic phenomena such as semantic roles, constituent structure position or relative syntactic rank. The inventory of grammatical functions is assumed to be universal, and each grammatical function is supposed to be associated with a distinct pattern of syntactic behaviour. The optimal inventory and the syntactic status of its members are based on three generalizations: (1) the functional hierarchy, which determines constraints on anaphoric binding and semantic role mapping; (2) the classification of grammatical functions into governable vs. ungovernable, semantically restricted vs. unrestricted GFs and terms vs. non-terms, as well as the related cross-classification of GFs in lexical mapping theory; (3) individual properties of specific grammatical functions, primarily subjects. This defines the core five-way distinction between SUBJ, OBJ, OBJ_θ, OBL_θ

and ADJ. Four grammatical functions – POSS, COMP, XCOMP and PREDLINK – stand somewhat apart due to being uniquely associated with very specific argument types: nominal possessors, clausal complements and nonverbal predicates. This has resulted in attempts to eliminate PREDLINK and assimilate POSS to SUBJ and (X)COMP to non-clausal other grammatical functions, but there are compelling independent arguments in favour of preserving their distinct status. In addition to these GFs, LFG makes use of so-called overlay functions, which represent positions additionally occupied by GFs that are required for cross-clausal or discourse continuity.

This approach goes against the general trend in formal syntax but aligns LFG very well with typological and functional approaches to language, where *grammatical relations* are direct counterparts to the LFG *grammatical functions*. In spite of the superficial similarity, however, there is a crucial difference between the two approaches: typology does not generally assume one specific system of grammatical relations to be universal, while LFG is concerned with universality, at least in theory. This focus on universality implies that the LFG notions of grammatical functions are quite removed from their traditional definitions. In particular, there have been interesting developments in the treatment of subjects: Manning (1996) replaces subject and object with more abstract functions PIVOT and CORE that receive an inverse mapping in ergative languages, while Falk (2006) retains the traditional SUBJ as the most prominent argument ($\widehat{\text{GF}}$) while adding the overlay function PIVOT to account for those subjecthood properties that can be associated with other arguments in syntactically non-accusative languages. The distinction between OBJ and OBJ$_\theta$ has also been extended beyond its traditional understanding, with OBJ$_\theta$ being used for unmarked direct objects in differential object marking languages (Dalrymple & Nikolaeva 2011) and for coindexed arguments in polysynthetic languages (see Lander et al. 2021 for a rather radical approach). Finally, the LFG use of a distinct GF (X)COMP for clausal complements is unique in theoretical and typological literature and allows a wide range of intriguing generalizations.

Grammatical functions are a cornerstone of LFG, and their analysis is in line with the general spirit of this framework, which avoids reductionism to the extent of sometimes being overtly redundant in splitting linguistic phenomena into several mechanisms operating at different levels. The framework itself puts no constraint on the relationship between these levels; determining to what extent the mapping is regular becomes an empirical question. There is no formal obstacle to eliminating grammatical functions from LFG if it can be demonstrated that they can be reduced to other mechanisms. However, all such attempts to date

Oleg Belyaev

have been unsuccessful, which demonstrates the viability of the LFG approach to grammatical functions.

Acknowledgments

Abbreviations

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

ADD	additive	HRSY	hearsay evidential (Meskwaki)
AOR	aorist	INTERJ	interjection
AUG	augmentative (Hakha Lai)	IN	inessive
AV	active voice (Tagalog)	IO	indirect object
BV	benefactive voice (Tagalog)	IV	instrumental voice (Tagalog)
CL	clitic	LAT	lative
CL1	first agreement class (East Caucasian languages)	LNK	linker
		MAL	malefactive
CL2	second agreement class (East Caucasian languages)	OV	objective voice (Tagalog)
		PCVB	participle-converb (Ashti)
CONN	connective (Hakha Lai)	POS	positive
COMPL	completive	POT	potential
DIM	diminutive	PQP	pluperfect (Moksha Mordvin)
DIR	directive		
DV	dative/locative voice (Tagalog)	PRET	preterite
		SUPER	location above landmark
EL	elative		
EMPH	emphatic	TMPR	temporal (Moksha Mordvin)
FV	final vowel (Kichaga)		

References

Alsina, Alex. 1993. *Predicate composition: A theory of syntactic function alterna-tions*. Stanford: Stanford University. (Doctoral dissertation).

Alsina, Alex. 2008. A theory of structure-sharing: Focusing on long-distance de-pendencies and parasitic gaps. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '08 conference*. Stanford: CSLI Publications.

Alsina, Alex, K. P. Mohanan & Tara Mohanan. 2005. How to get rid of the COMP. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '05 con-ference*, 21–41. Stanford: CSLI Publications.

Alsina, Alex, Tara Mohanan & K. P. Mohanan. 1996. Untitled submission to the lfg list. 3 sept 1996.

Alsina, Alex & Eugenio M. Vigo. 2014. Copular inversion and non-subject agree-ment. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 5–25. Stanford: CSLI Publications.

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. 2007. The major functions of the noun phrase. In *Language typology and syntactic description: Complex constructions*, vol. 1, 62–154. Cam-bridge, UK: Cambridge University Press.

Andrews, Avery D. Forthcoming. Clause structure and configurationality. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Lan-guage Science Press.

Asher, Nicholas. 1993. *Reference to abstract objects in discourse* (Studies in Lin-guistics and Philosophy 50). Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-011-1715-9.

Asudeh, Ash. 2012. *The logic of pronominal resumption* (Oxford Studies in The-oretical Linguistics). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199206421.001.0001.

Asudeh, Ash. Forthcoming. Glue Semantics. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Belyaev, Oleg. 2015. Systematic mismatches: Coordination and subordination at three levels of grammar. *Journal of Linguistics* 51(2). 267–326.

Belyaev, Oleg. Forthcoming(a). Core concepts of LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Belyaev, Oleg. Forthcoming(b). Grammatical functions: their classification and properties. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Oleg Belyaev

Belyaev, Oleg. Forthcoming(c). Introduction to LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Belyaev, Oleg, Anastasia Kozhemyakina & Natalia Serdobolskaya. 2017. In defense of COMP: Complementation in Moksha Mordvin. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '17 conference*, 83–103. Stanford: CSLI Publications.

Berman, Judith. 1999. Does German satisfy the Subject Condition? In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '99 conference*. Stanford: CSLI Publications.

Berman, Judith. 2003. *Clausal syntax of German*. Stanford: CSLI Publications.

Berman, Judith. 2007. Functional identification of complement clauses in German and the specification of COMP. 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*, 69–83. Stanford: CSLI Publications.

Bickel, Balthasar. 2010. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199281251.013.0020.

Börjars, Kersti & Nigel Vincent. 2008. Objects and OBJ. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '08 conference*, 150–168. Stanford: CSLI Publications.

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 & Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63(4). 741–782. DOI: 10.2307/415717.

Bresnan, Joan & Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21. 147–186.

Butt, Miriam. Forthcoming. Case. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A grammar writer's cookbook*. Stanford: CSLI Publications.

Chisarik, Erika & John R. Payne. 2003. Modelling possessor constructions in LFG: English and Hungarian. In Miriam Butt & Tracy Holloway King (eds.), *Nominals: Inside and out*, 181–199. Stanford: CSLI Publications.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding* (Linguistic Inquiry Monographs). Cambridge, MA: The MIT Press.

Creissels, Denis. 2019. Grammatical relations in Mandinka. In Alena Witzlack-Makarevich & Balthasar Bickel (eds.), *Argument Selectors: A new perspective on grammatical relations* (Typological Studies in Language 123), 301–348. Amsterdam: John Benjamins. DOI: 10.1075/tsl.123.09cre.

Culicover, Peter W. & Ray Jackendoff. 1997. Semantic subordination despite syntactic coordination. *Linguistic Inquiry* 28(2).

Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199271092.001.0001.

Dahlstrom, Amy. 2009. OBJθ without OBJ. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '09 conference*, 222–239. Stanford: CSLI Publications.

Dahlstrom, Amy. 2014. Multiple oblique arguments in Meskwaki. In J. Randolph Valentine & Monica Macaulay (eds.), *Papers of the forty-second Algonquian conference 2010*, 56–68. Albany, NY: State University of New York Press.

Dalrymple, Mary. 1993. *The syntax of anaphoric binding*. Stanford: CSLI Publications.

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.

Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. Stanford: CSLI Publications.

Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '00 conference*, 104–121. Stanford: CSLI Publications.

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.

Dixon, R. M. W. 1979. Ergativity. *Language* 55(1). 59–138. DOI: 10.2307/412519.

Oleg Belyaev

Dixon, R. M. W. 1994. *Ergativity*. Cambridge/New York: Cambridge University Press. DOI: 10.1017/cbo9780511611896.

Dowty, David R. 1982. Grammatical relations and Montague Grammar. In Pauline Jacobson & Geoffrey K. Pullum (eds.), *The nature of syntactic representation*, 79–130. Dordrecht: Reidel. DOI: 10.1007/978-94-009-7707-5_4.

Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67. 547–619. DOI: 10.2307/415037.

Falk, Yehuda N. 2001. *Lexical-Functional Grammar: An introduction to parallel constraint-based syntax*. Stanford: CSLI Publications.

Falk, Yehuda N. 2006. *Subjects and universal grammar: An explanatory theory*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9780511486265.

Fassi Fehri, Abdelkader. 1988. Agreement in Arabic, binding and coherence. In Michael Barlow & Charles A. Ferguson (eds.), *Agreement in natural language*, 107–158. Stanford: CSLI Publications.

Fillmore, Charles J. 1968. The case for case. In Emmon W. Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 1–88. New York: Holt, Rinehart & Winston.

Findlay, Jamie Y. & Anna Kibort. Forthcoming. Argument structure and mapping theory. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Forst, Martin. 2006. COMP in (parallel) grammar writing. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '06 conference*, 222–239. Stanford: CSLI Publications.

Frantz, Donald G. 1981. *Grammatical relations in Universal Grammar*. Bloomington, IN: Indiana University Linguistics Club.

Ganenkov, Dmitry. 2013. Diachrony of experiencer subject marking: evidence from East Caucasian. In *The diachronic typology of non-canonical subjects* (Studies in Language Companion Series 140), 231–2. Amsterdam: John Benjamins.

Haspelmath, Martin. 2009. Framework-free grammatical theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 341–365. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199677078.013.0014.

Haug, Dag. 2013. Partial control and anaphoric control in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '13 conference*, 274–294. Stanford: CSLI Publications.

Haug, Dag. 2014. Partial dynamic semantics for anaphora: Compositionality without syntactic coindexation. *Journal of Semantics* 31(4). 457–511. DOI: 10.1093/jos/fft008.

Haug, Dag & Tatiana Nikitina. 2012. The many cases of non-finite subjects: The challenge of "dominant" participles. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '12 conference*, 292–311. Stanford: CSLI Publications.

Haug, Dag & Tatiana Nikitina. 2015. Feature sharing in agreement. *Natural Language & Linguistic Theory* 34. 865–910. DOI: 10.1007/s11049-015-9321-9.

Hellan, Lars. 1988. *Anaphora in Norwegian and the theory of grammar*. Dordrecht: Walter de Gruyter. DOI: 10.1515/9783110849134.

Hualde, José Ignacio, Bernard Oyharçabal & Jon Ortiz de Urbina. 2003. Verbs. In José Ignacio Hualde (ed.), *A grammar of Basque* (Mouton Grammar Library 26), 195–322. Berlin: Mouton de Gruyter.

Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: Massachusetts Institute of Technology. (Doctoral dissertation).

Johnson, David E. 1977. On relational constraints on grammars. In Peter Cole & Jerrold M. Sadock (eds.), *Grammatical relations* (Syntax and Semantics 8). New York: Academic Press.

Johnson, David E. & Paul M. Postal. 1981. *Arc Pair Grammar*. Princeton: Princeton University Press. DOI: 10.1515/9781400855551.

Jónsson, Jóhannes Gísli. 2003. Not so quirky: On subject case in Icelandic. In E. Brandner & H. Zinsmeister (eds.), *New perspectives on case and case theory*, 129–164. Stanford: CSLI Publications.

Kaplan, Ronald M. 2017. Preserving grammatical functions in LFG. In Victoria Rosén & Koenraad De Smedt (eds.), *The very model of a modern linguist – In honor of Helge Dyvik* (Bergen Language and Linguistics Studies 8), 127–142. Bergen: Bergen Language & Linguistics Studies (BeLLS). DOI: 10.15845/bells.v8i1.1342.

Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. Cambridge, MA: The MIT Press. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 29–130).

Kaplan, Ronald M. & Annie Zaenen. 1989. Functional precedence and constituent structure. In Chu-Ren Huang & Keh-Jiann Chen (eds.), *Proceedings of ROCLING II*, 19–40. Taipei.

Kaplan, Ronald M. & Annie Zaenen. Forthcoming. Long-distance dependencies. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.

Oleg Belyaev

Kibort, Anna. 2006. On three different types of subjectlessness and how to model them in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '06 conference*, 289–309. Stanford: CSLI Publications.

Kibort, Anna. 2014. Mapping out a construction inventory with (Lexical) Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 262–282. Stanford: CSLI Publications.

Kibrik, Alexander E. 2000. K probleme jadernyx aktantov i ix "nekanoničeskogo kodirovanija": svidetel'stva arčinskogo jazyka [towards the problem of core arguments and their "non-canonical marking": evidence from archi]. *Voprosy Jazykoznanija* 5. 32–67.

King, Tracy Holloway. 1995. *Configuring topic and focus in Russian*. Stanford: CSLI Publications.

King, Tracy Holloway & Annie Zaenen. 2004. F-structures, information structure, and discourse structure. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '04 conference*. Stanford: CSLI Publications.

Kiparsky, Paul & Carol Kiparsky. 1970. Fact. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics*, 143–173. The Hague: Mouton.

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

Laczkó, Tibor. 2004. Grammatical functions, LMT, and control in the Hungarian DP revisited. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '04 conference*, 313–333. Stanford: CSLI Publications.

Laczkó, Tibor. 2009. On the *-Ás* suffix: Word formation in the syntax? *Acta Linguistica Hungarica* 56(1). 23–114. DOI: 10.1556/aling.56.2009.1.2.

Laczkó, Tibor. 2017. Modelling (in)definiteness, external possessors and (typological) variation in Hungarian possessive DPs. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '17 conference*, 243–263. Stanford: CSLI Publications.

Laczkó, Tibor & György Rákosi. 2019. Pronominal possessors and syntactic functions in the Hungarian possessive noun phrase. In Miriam Butt, Tracy Holloway King & Ida Toivonen (eds.), *Proceedings of the LFG '19 conference*, 149–169. Stanford: CSLI Publications.

Lander, Yury, Oleg Belyaev & Irina Bagirokova. 2021. (Almost) everything is oblique in West Circassian. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG '21 conference*, 223–242. Stanford: CSLI Publications.

Legate, Julie Anne. 2012. Types of ergativity. *Lingua*. DOI: 10.1016/j.lingua.2011.10.014.

Letuchiy, Alexander. 2016. Agreement with complement clauses in Adyghe. *RHEMA* 4. 83–104.

Letuchiy, Alexander B. 2012. O nekotoryx svojstvax sentencial'nyx aktantov v russkom jazyke [On certain features of complement clauses in Russian]. *Voprosy jazykoznanija* 5. 57–87.

Lovestrand, Joseph. 2009. French causatives in LFG: Adverbial complements as evidence of subjecthood properties of the causee. *GIALens.* 2.

Lowe, John J., Adriana Molina-Muñoz & Antonia Ruppel. 2021. Raising and passive in Sanskrit. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG '21 conference*, 243–263. https://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG-2021/lfg2021-lmr.pdf.

Manning, Christopher D. 1992. *Romance is so complex.* Tech. rep. CSLI-92-168. Stanford: Stanford University. http://nlp.stanford.edu/~manning/papers/romance.ps.

Manning, Christopher D. 1996. *Ergativity: Argument structure and grammatical relations* (Dissertations in Linguistics). Stanford: CSLI Publications.

Melchin, Paul. 2019. *The semantic basis for selectional restrictions.* University of Ottawa. (Doctoral dissertation).

Melvold, Janis. 1991. Factivity and definiteness. *MIT Working Papers in Linguistics.*

Mohanan, K. P. 1982. Grammatical relations and clause structure in Malayalam. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 504–589. Cambridge, MA: The MIT Press.

Moravcsik, Edith A. 1978. Agreement. In Joseph H. Greenberg (ed.), *Universals of human language*, vol. 4, 331–374. Stanford: Stanford University Press.

Müller, Stefan & António Branco. 2021. Anaphoric binding. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*. Berlin: Language Science Press.

Needham, Stephanie & Ida Toivonen. 2011. Derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '11 conference*, 401–421. Stanford: CSLI Publications.

Patejuk, Agnieszka. Forthcoming. Coordination. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Patejuk, Agnieszka & Adam Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the joint 2016 conference on Head-Driven Phrase Structure Grammar and Lexical Functional Grammar*, 541–559. Stanford: CSLI Publications.

Perlmutter, David M. 1971. *Deep and surface constraints in syntax* (The Transatlantic Series in Linguistics). New York: Holt, Rinehart & Winston.

Oleg Belyaev

Peterson, David A. 2007. *Applicative constructions* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.

Peterson, Philip L. 1997. *Fact, Proposition, Event* (Studies in Linguistics and Philosophy 66). Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-015-8959-8.

Polinsky, Maria & Eric Potsdam. 2002. Backward control. *Linguistic Inquiry* 33. 245–282. DOI: 10.1162/002438902317406713.

Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics.* Stanford: CSLI Publications.

Potsdam, Eric & Maria Polinsky. 2012. Backward raising. *Syntax* 12. 75–108.

Przepiórkowski, Adam. Forthcoming. LFG and HPSG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar.* Berlin: Language Science Press.

Rákosi, György. Forthcoming. Anaphora. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar.* Berlin: Language Science Press.

Rhodes, Richard R. 1994. Agency, inversion, and thematic alignment in Ojibwe. In *Proceedings of the 20th annual meeting of the Berkeley Linguistics Society*, 431–446. DOI: 10.3765/bls.v20i1.1436.

Rizzi, Luigi. 1990. *Relativized Minimality.* Cambridge, MA: The MIT Press.

Ross, John R. 1967. *Constraints on variables in syntax.* Cambridge, MA: Massachusetts Institute of Technology. (Doctoral dissertation).

Schwarze, Christoph. 2001. On the representation of French and Italian clitics. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '01 conference*, 280–304. Stanford: CSLI Publications.

Sells, Peter. 2006. Using subsumption rather than equality in functional control. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '06 conference*, 455–473. Stanford: CSLI Publications.

Sells, Peter. Forthcoming. LFG and Minimalism. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar.* Berlin: Language Science Press.

Siewerska, Anna. 2013. Verbal person marking. In Matthew Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online.* Chapter 102. Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/102.

Snijders, Liselotte. 2015. *The nature of configurationality in LFG.* Oxford: University of Oxford. (D.Phil. Thesis).

Sulger, Sebastian. 2015. *Modeling nominal predications in Hindi/Urdu.* University of Konstanz. (Doctoral dissertation).

Van Valin, Robert, Jr. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning, and function.* Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9781139166799.

Varaschin, Giuseppe. Forthcoming. LFG and Simpler Syntax. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Vincent, Nigel. Forthcoming. Raising and control. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Yates, Nicholas. 2002. French causatives: A biclausal account in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '02 conference*, 390–407. Stanford: CSLI Publications.

Zaenen, Annie. 1985. *Extraction rules in Icelandic*. New York: Garland.

Zaenen, Annie. Forthcoming. Information structure. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Zaenen, Annie & Richard Crouch. 2009. OBLs hobble computations. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '09 conference*, 644–654. Stanford: CSLI Publications.

Zaenen, Annie & Elisabet Engdahl. 1994. Descriptive and theoretical syntax in the lexicon. In Beryl T. S. Atkins & Antonio Zampolli (eds.), *Computational approaches to the lexicon*, 181–212. Oxford: Clarendon Press.

Chapter 2

Argument structure and mapping theory

Jamie Y. Findlay
University of Oslo

Roxanne Taylor
University of Huddersfield

Anna Kibort
Trinity College Library

This chapter presents the LFG view of two closely related areas of inquiry: argument structure, a level of structure which represents the syntactically realisable arguments of a predicate, and mapping theory, the theory of how those arguments are linked to grammatical functions at f-structure, as well as of alternations in this linking brought about by processes like passivisation. After introducing some preliminary concepts, the chapter explores various approaches within LFG: the earliest work using lexical rules to explain argument alternations, the "classical" version of Lexical Mapping Theory (LMT) developed in the late '80s and early '90s, and various subsequent modifications, extensions, and re-imaginings of LMT, including contemporary work focussing on the formal status of argument structure and mapping theory, and their connection to the rest of the grammar.

1 Introduction

Predicates have both syntactic and semantic arguments, and the two are not always aligned. For instance, expletives, as shown in (1), are syntactic but not semantic arguments of their verbal governors:

(1)   a.   *It* is snowing.

     b.   *There* seems to be a problem.

On the other hand, there are verbs like *saddle*: conceptually, a saddling event involves three entities, the saddler, the saddled (usually a horse), and the saddle itself, but only the first two are expressible in the syntax (cf. Bresnan 1980). Similarly, there are clear patterns regarding which kinds of semantic arguments are realised by which kinds of grammatical functions – in general, more Agent-like arguments are more likely to be subjects than more Patient-like arguments, which are more likely to be objects – but there are also exceptions. There are, for example, verbs which seem to express the same type of event but to realise the semantic participants differently in the syntax (Rappaport 2006 [1983]: 132):

(2) a. Fred fears the prospect of failure.

　　 b. The prospect of failure frightens Fred.

(3) a. I like a job well done.

　　 b. A job well done pleases me.

Due to these kinds of mismatches, neither syntactic nor semantic arguments can be reduced to the other, and instead we need some intervening level of representation that can mediate the relationship between them. This is what is known as ARGUMENT STRUCTURE, and in LFG is often taken to constitute a separate module of the grammar called a-structure. Although it sits between syntax and (lexical) semantics, argument structure is often taken to itself be a specifically syntactic level of representation (Alsina 2001),[1] whose primary purpose is to explain a predicate's syntactic valency patterns – while acknowledging that at least some of these explanations are to be found in lexical semantic properties. The arguments represented at argument structure are therefore those which can or must be realised syntactically.

Explaining how exactly these arguments are realised is the purview of MAPPING THEORY. Such a theory seeks generalisations in the mapping between argument structure and syntax proper, and to explain any alternations which are possible (such as passivisation, causativisation, detransitivisation, etc.). In LFG, this means determining what GRAMMATICAL FUNCTION (GF) the argument will instantiate – overt phrasal realisation is then handled by the language-specific phrase-structure rules or case-marking system which determines how particular GFs surface (see Belyaev forthcoming(a),(b) [this volume] for more on LFG's view of grammatical functions and their relation to phrasal syntax).

---

[1]Indeed, in the Minimalist tradition, argument structure is often represented in the phrasal syntax itself – see Harley (2011) and references therein for an overview.

As part of this LFG handbook, the present chapter focuses on providing a survey of work on argument structure and mapping theory which takes a Lexical-Functional approach.[2] The structure of the chapter is as follows: we begin, in Section 2, with a brief high-level introduction to some of the questions and phenomena which we will return to throughout the chapter. Section 3 then looks at the earliest work on these problems in LFG, which used LEXICAL RULES to account for argument alternations. Section 4, the largest of the chapter, presents the still-canonical theory of argument structure and mapping developed in the late 1980s and early 1990s, known as LEXICAL MAPPING THEORY (LMT). Section 5 examines a different version of LMT, that of Kibort (2007, *i.a.*) which, among other things, is designed to extend the empirical coverage of the mapping theory to so-called morphosemantic alternations. Section 6 delves more deeply into some formal issues and alternative proposals, before Section 7 concludes.

2 Background and basic concepts

2.1 From semantics to syntax

There are regularities in the ways that semantic participants of predicates are realised syntactically. For example, in a nominative-accusative language like English, when a verb describes an event that has a volitional Agent and a Theme or Patient affected by the event, the Agent will be realised as the active voice subject and the Theme/Patient as the object:

(4) a. Your dog is chasing my rabbit!
     (cf. # My rabbit is chasing your dog![3])

   b. The engineers will build the bridge there.
     (cf. # The bridge will build the engineers there.)

   c. The teacher opened the cupboard.
     (cf. # The cupboard opened the teacher.)

---

[2]For general introductions as well as critical overviews of work in other traditions, the reader is directed to Grimshaw (1990), Comrie (1993), Levin & Rappaport Hovav (2005), Ramchand (2014), Williams (2015); and for a different perspective on the LFG literature, see Dalrymple et al. (2019: ch. 9).

[3]The point of these anomalous alternatives is to illustrate that the (prototypical) situations presented are expressed via the (a) encodings, where the Agent is a subject and the Theme/Patient an object, rather than the *a priori* equally plausible (b) encodings, where the pairings of semantic and syntactic roles are reversed. The (b) sentences are of course perfectly grammatical strings of Engilsh, but they describe situations which are at odds with our real-world knowledge or expectations, precisely because the subjects in each case are interpreted as Agents.

Similarly, if the sentence expresses an Instrument used to perform the action described, along with the Theme/Patient, then the Instrument is the subject and the Theme/Patient the object:

(5)   a.   The key opened the cupboard.
           (cf. # The cupboard opened the key.)

But if the Agent is also included, then *it* is the subject:

(6)   The teacher opened the cupboard with the key.

This generalisation goes back to Fillmore (1968: 33), who expresses it as follows:

(7)   If there is an A [= Agent], it becomes the subject; otherwise, if there is an
      I [= Instrument], it becomes the subject; otherwise, the subject is the O
      [= objective, i.e. Theme/Patient].

This is a productive rule (at least in English), as can be seen from the fact that invented words will also follow the same pattern. Alsina (1996: 5–6), for instance, imagines a verb *obliquate*, meaning 'build or place in an oblique position or direction', and notes the clear intuition that, if such a verb existed, we would say things like (8a), but not like (8b):

(8)   a.   Jim obliquated the door of the closet.
      b.   # The door of the closet obliquated Jim.

   All this goes to illustrate a key explanandum: the semantic relationship which an argument bears to its verb is also implicated in determining its syntactic relationship, but in what way precisely? Mapping theory is interested in discovering the nature of this connection, and in finding generalisations over the links between semantic and syntactic relationships.

   The observation in (7) induces a ranking of semantic/thematic roles,[4] where the highest available argument becomes the subject:

(9)   A > I > O

This can be seen as a precursor to the well-known THEMATIC HIERARCHY (Jackendoff 1972: 43), of which there have been many versions. The one which has been most influential in LFG comes from Bresnan & Kanerva (1989: 23), and is shown in (10):

---

[4]We will use these two terms interchangeably in this chapter, drawing no theoretical distinction between them.

(10)   *The Thematic Hierarchy:*
       Agent > Beneficiary > Recipient/Experiencer
                                    > Instrument > Theme/Patient > Location

Arguments which are more thematically "prominent" on this hierarchy tend to be realised by more grammatically "prominent" GFs, e.g. as defined by the Keenan-Comrie hierarchy (Keenan & Comrie 1977; see also Belyaev forthcoming(b) [this volume]) – in particular, the subj function is usually taken by the the argument highest on the thematic hierarchy (Grimshaw 1990, Speas 1990). This insight is often at the core of mapping theories, and so the thematic hierarchy figures centrally in the standard version of Lexical Mapping Theory, which we explore in Section 4, as well as in other approaches discussed below.

The use of thematic hierarchies has also been challenged, however. For one thing, a consistent list of roles and definitions has proved elusive, and classification of arguments can therefore be problematic and open to disagreement (Gawron 1983, Dowty 1991, Ackerman & Moore 2001, Davis 2011). For another, even when a set of roles is agreed on, the question of their relative ordering has not been settled, and different hierarchies have been proposed for different phenomena, or even for the same phenomenon (Newmeyer 2002: 65ff. Levin & Rappaport Hovav 2005: ch. 6; Rappaport Hovav & Levin 2007). While it is clearly possible that different orderings could be relevant for different things, the extent of the variability in the literature, even with respect to one and the same phenomenon, stands in stark contrast to the putative appeal of a unifying thematic hierarchy where a fixed set of roles is used in order to abstract away from predicate-specific semantic entailments. Because of these concerns, some recent work in LFG's mapping theory, most notably that of Kibort (2007, *i.a.*), has attempted to do without thematic roles altogether. We discuss Kibort's work in Section 5.

Some questions of mapping depend not on the semantic relationship between an argument and its verb, but rather on lexical semantic properties of the verb itself. For example, *break* and *hit* both take Agent and Patient arguments, but *break* has an intransitive alternant, where the Patient appears as the subject, which is impossible with *hit*:

(11)   a.   The teacher broke the ruler.
       b.   The ruler broke.

(12)   a.   The teacher hit the ruler.
       b.   * The ruler hit.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Fillmore (1970) observes that this contrast is not a lexical idiosyncrasy of these two verbs, but actually applies to two large classes of semantically-related verbs, as shown in (13–14):

(13) a. The teacher {bent / folded / shattered / cracked / ...} the ruler.

  b. The ruler {bent / folded / shattered / cracked / ...}.

(14) a. The teacher {slapped / struck / bumped / stroked / ...} the ruler.

  b. * The ruler {slapped / struck / bumped / stroked / ...}.

Once again, we can see that this is a productive generalisation if we examine our intuitions about invented forms. For example, let us imagine a verb *jellate*, meaning 'to turn to jelly'. It is clear that this verb could appear in the same constructions as *break*.

(15) a. The wizard jellated the box.

  b. The box jellated.

But if we invent a word like *coude*, meaning 'to touch with one's elbow', it is just as clear that it will pattern with *hit*:

(16) a. I couded the wall.

  b. * The wall couded.

We do not want to simply stipulate the possibilities for each new verb, since then we fail to capture the regularity and productivity of our intuitions.

A mapping theory ought to give an account of these patterns. To do this, it must have access to detailed lexical semantic information, such as event structure. For example, a hitting event does not necessarily result in a change of state in the affected entity, whereas a breaking event does; that is, the structure of a hitting event does not contain a result state, in Ramchand's (2008) terms. Now, this may be expressed in the semantic role assigned to the affected entity – in some theories, the difference between Patient and Theme is that the former undergoes a change of state while the latter does not. But often such nuances are not captured by a simple semantic role analysis – for example, the thematic hierarchy in (10) collapses Theme and Patient into a single position – and it is certainly not apparent that there are any principled limits on what kinds of lexical semantic information can be relevant for questions of mapping, so it is quite possible that mapping theory needs access to a very rich representation of lexical semantics. In general, argument structure proposals in LFG have not taken up this challenge,

instead treating this level of representation as relatively informationally impoverished (it is often no more than a list of arguments and their associated thematic roles). Nevertheless, there have been, and continue to be a growing number of, exceptions, which we examine in Section 6.1.

2.2 Argument alternations

Accounting for the syntactic realisation of semantic arguments means also addressing the fact that a single predicate may permit multiple ways of expressing its arguments (including not expressing some of them at all) – that is, the existence of ARGUMENT ALTERNATIONS, such as that between the transitive and the inchoative illustrated in (11), above. Perhaps the most famous and well-studied of these is the active-passive alternation, a typologically common pattern whereby a transitive verb alternates with an intransitive in which the subject argument of the transitive form is either unexpressed or expressed as a non-core, oblique grammatical function instead:

(17)   a. *Active:*
           The dog chased the rabbit.
        b. *Passive:*
           The rabbit was chased (by the dog).

One important property of the active-passive alternation is that it does not involve any change in lexical semantics. That is, the situations described by (17a) and (the long version of) (17b) are truth-conditionally equivalent, and so this alternation is described as MEANING-PRESERVING (cf. Sadler & Spencer 1998). This label is slightly infelicitous, however, since once we look beyond mere truth conditions there are of course changes to other aspects of "meaning", writ large: for instance, the information-structure Topic is the dog in (17a) but the rabbit in (17b) . This is not at all surprising, however, since language abhors true synonymy (Cruse 1986, Goldberg 2019), and variation of whatever kind is inevitably operationalised for communicative purposes (Clark 1987, Eckert 2018) – nevertheless, it does mean that the term "meaning-preserving" must be understood in a suitably narrow sense.

Such meaning-preserving alternations are known as MORPHOSYNTACTIC, since they are morphological operations which alter the syntactic alignment of participants; this is in contradistinction to MORPHOSEMANTIC alternations, which involve changes in (truth-conditional) lexical meaning. Another example of a morphosyntactic alternation is locative inversion, illustrated in (18) for Chicheŵa

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

(Bresnan & Kanerva 1989: 2). In this alternation (also found in English, as indicated by the translations below – see Bresnan 1994), a locative phrase which normally appears as an oblique can surface as a subject, demoting the subject of the non-inverted form to object:[5]

(18) a. Chi-tsîme chi-li ku-mu-dzi.
          7-well   7-be   17-3-village
          'The well is in the village.'

     b. Ku-mu-dzi ku-li chi-tsîme.
          17-3-village 17-be 7-well
          'In the village is a well.'

Once again, this affects certain properties of a sentence's information structure, for instance changing what is available for contrastive focus (Bresnan & Kanerva 1989: 35, Bresnan 1994: 86–87), but it does not alter the truth-conditional meaning.

Morphosemantic alternations, on the other hand, change the lexical meaning of a predicate – a change which may then have syntactic effects, though these are in a sense only incidental, merely following as automatic consequences of the lexical semantic changes (Kibort 2004: 374). Examples include many of the alternations listed in Levin (1993), such as the *spray/load* alternation shown in (19) or the dative shift alternation shown in (20):

(19) a. Carly loaded the wagon with barrels.

     b. Carly loaded barrels onto the wagon.

(20) a. Julian brought Elim the message.

     b. Julian brought the message to Elim.

In (19a), the Goal/Location *the wagon* is realised as the object, and in this case there is a "holistic" interpretation (Levin 1993: 50), whereby the Goal/Location is understood to be fully affected by the action (i.e. the wagon is filled up with barrels). This entailment is absent from the sentence in (19b), where the Theme is realised as the object instead. Similarly, in (20a), there is an entailment that the dative-shifted Goal object is animate (Goldberg 1995: 146–147), but this same constraint does not hold of the Goal argument in the prepositional variant (20b), as illustrated by the following contrast:

---

[5]Numbers indicate noun classes: this is in part how we can tell that the locative is the subject in (18b), since the verb now agrees with it in this respect.

(21) a. # Julian brought Elim's study the message.

     b. Julian brought the message to Elim's study.

Both of these alternations involve differing syntactic realisations of the same arguments, but unlike the morphosyntactic alternations shown above, they also change certain properties of the truth-conditional meanings expressed by their governing verbs. Other morphosemantic alternations, such as the causative, also introduce *new* arguments, rather than simply rearranging existing arguments. The causative introduces a new Causer argument, which brings about the event described by the predicate. Here is a classic example from Turkish (Comrie 1974: 5):

(22) a. Hasan öl-dü.
        Hasan die-PST
        'Hasan died.'

     b. Ali Hasan-ı   öl-dür-dü.
        Ali Hasan-OBJ die-CAUS-PST
        'Ali killed Hasan.' (lit. 'Ali made Hasan die.')

As can be seen, this also has syntactic effects, since causativisation increases the valency of the predicate. Here an intransitive becomes a transitive, and the previous subject is demoted to object.

The world's languages are replete with a wide and varied selection of argument alternations, both meaning-preserving and meaning-altering, many of which are highly productive. Any mapping theory must therefore be capable of giving an account of such alternations in general, and this has been a major focus of research, as we will see below.

3 Lexical rules

Argument alternations have been at the heart of work in LFG since the very beginning. The seeds of LFG as a framework can be found in Bresnan's (1978) work on the psychological plausibility of transformational grammars, illustrating how the passive can be profitably viewed as an operation on lexical representations, rather than on phrase-level syntactic structures. Bresnan (1980) presents this analysis in a more recognisably LFG-like form, and extends the approach to the formation of intransitives and middles in English. In this and much other early work in LFG, argument alternations are treated as involving LEXICAL RULES,

which systematically relate the different alternants of the same verb (e.g. active and passive). In this section, we give a brief overview of this approach, and highlight some of the reasons why it has fallen out of favour in recent work.

In Bresnan (1980), lexical items are assumed to possess abstract predicate-argument structures, which characterise "those arguments of a semantic predicate that are open to grammatical interpretation" (Bresnan 1980: 100). Such argument positions are then associated with grammatical functions by various (undiscussed) lexical processes, with the result being a LEXICAL FORM – recognisable as what would become in LFG the SEMANTIC FORM value of a PRED attribute (Belyaev forthcoming[a] [this volume]). For example, the lexical form for transitive *read*, as in *John read my letter*, is given in (23) (Bresnan 1980: 116):

(23)   *read*   ⟨   (SUBJ)   (OBJ)   ⟩

Here the first argument, corresponding to the reader, is linked to SUBJ, and the second argument, the thing read, is linked to OBJ. The exact nature of this initial linking of arguments to GFs is not spelled out explicitly, and is generally taken to follow from some intrinsic pairings of roles and syntactic functions. What is more, in this early work, the specific role of each argument is not labelled in the representation, and must be inferred from the combination of their ordering and lexical idiosyncrasies of meaning. In other work (e.g. Baker 2006 [1983]), lexical forms are shown with semantic roles alongside their associated GFs, thus highlighting both sides of the linking question explicitly in the representation. For the sake of clarity, we will follow this convention for the rest of this section; thus instead of (23), we will write (24) for the lexical form of *read*:

(24)   *read*   ⟨   Agent   Theme   ⟩
                        (SUBJ)   (OBJ)

However such structures are represented, once the links between arguments and GFs are in place, other rules can then apply to manipulate them, capturing the effect of various argument alternations. For example, intransitivisation is achieved by the following lexical rule (Bresnan 1980: 116):

(25)   *Intransitivisation:*
        (OBJ) ↦ ∅

Here the argument previously linked to OBJ is instead assigned the special null GF ∅, which indicates that the argument is existentially bound in the semantics, and is not expressed overtly in the syntax. The application of (25) to (24) results in the lexical form in (26), corresponding to the intransitive form of *read*, as in *John read all night*.

(26)   *read*   ⟨   Agent   Theme   ⟩
                    (SUBJ)   ∅

It is clear to see how this approach can be extended to other, more complex alternations. Bresnan (1982), for instance, proposes the following lexical rule for passivisation:[6]

(27)   *Passivisation:*
       (SUBJ) ↦ ∅/(OBL$_{\text{AGENT}}$)
       (OBJ) ↦ (SUBJ)

This demotes the subject to either the unexpressed null GF (as in the English short, Agent-less passive), or an oblique (as in the English long, *by*-passive), and promotes the object to subject.

One important strength of such lexical rules is that they manipulate grammatical functions, rather than surface constituent structures; that is, (27) promotes the OBJ, rather than, say, moving the post-verbal NP to the specifier position of IP. This means that the same rule can be used across the languages of the world, with language-specific variations falling out from the rules for c- to f-structure mapping in those languages (Belyaev forthcoming[a] [this volume]).[7] Such an approach is a corollary of the claim that argument alternations operate at the level of argument structure, and not directly on the phrasal syntax.

Lexical rules in LFG are taken to be REDUNDANCY RULES (Bresnan 1990: 638): they are not applied on-line in the process of parsing, but instead describe regular relations between items in the lexicon. In other words, the existence of a lexical form like (28a) implies the existence of a corresponding passive form like (28b), because of the existence of rule (27):[8]

(28)   a.   *read*   ⟨   Agent   Theme   ⟩
                          (SUBJ)   (OBJ)

       b.   *read*   ⟨   Agent   Theme   ⟩
                          (OBL$_{\text{AGENT}}$)   (SUBJ)

---

[6]Bresnan (1982) is in fact the *locus classicus* of the lexicalist approach to the passive in general. In the paper, Bresnan makes a compelling case against the prevailing wisdom that passivisation should be treated as a transformation, i.e. something that takes places in the phrasal syntax. Instead, she shows that it must be treated as a process occurring inside the lexicon. Bresnan et al. (2016: ch. 3) provide a contemporary presentation of the relevant arguments.

[7]This insight originates from work in Relational Grammar (e.g. Perlmutter & Postal 1977).

[8]Bresnan (1980, 1982) presents such rules as directional, so that the active maps to the passive, but they can also be seen as bidirectional, so that the existence of either kind of entry implies the other – this is how it is presented in Bresnan (1990), for example.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Such a restriction follows from Bresnan's (1980: 118) claim that "structures which are analyzed by lexical rules must be lexical structures, and cannot be syntactically derived". Bresnan (1982: 6) goes further, and proposes that alterations of argument-to-GF assignments can *only* take place in the lexicon, via lexical rules, and cannot be effected on-line by syntactic rules – she refers to this as the principle of DIRECT SYNTACTIC ENCODING. Although contemporary LFG makes much less (or no) use of lexical rules, it continues to maintain the first part of this principle, and treats all argument alternations as applying in the lexicon, not in the syntax.

While lexical forms, which appear at f-structure as the value of PRED attributes, are obtained by augmenting a predicate-argument structure with linkings to GFs, at this stage in the development of LFG the formal status of the predicate-argument structures themselves is not made explicit. They are certainly not a separate level of representation, akin to c- and f-structure (i.e. there is no a-structure). Indeed, it is not until Butt et al. (1997) that the formal position of argument structure in the LFG architecture is tackled head on – we will have more to say about this in Section 6.1.

A more urgent shortcoming of the early lexical rule approach is that there is no account of how the original assignment of GFs to arguments is accomplished – that is, as Falk (2001b: 96) observes, early LFG has a theory of *re*mapping, via lexical rules, but no theory of the initial mapping. Bresnan (1980: 112) briefly suggests some principles for default assignments, but this is not developed more fully. Since, as we observed in Section 2.1, the initial mapping is also amenable to systematic study, and exhibits a number of clear generalisations, this lacuna is therefore a significant one.

There is also the question of appropriately constraining lexical rules. Clearly the rule of intransitivisation given in (25) cannot apply freely to any verb with an object, otherwise we would expect examples like (29b) to be grammatical, contrary to fact:

(29) a. Naomi told the story to Jim.
      b. * Naomi told to Jim.

Lexical rules must be assigned syntactic, semantic, and morphological conditions in order to constrain their application. Even then, it remains a fact that lexical rules are very powerful formal devices: there are no in-principle constraints on what kinds of alternations can be described, which means that any remapping

can be represented, including some which are most unnatural in the world's languages (Bresnan 1980: 639ff.).[9]

The unconstrained expressive power of lexical rules arises from the fact that they are not MONOTONIC (Bresnan 1990): since such rules *overwrite* the original assignments of GFs to arguments, they are are not information-preserving.[10] Aside from the possibility of expressing unnatural alternations, another reason why non-monotonicity may be problematic is to do with processing. Arbitrary re-write rules render a system intractable (cf. Peters & Ritchie 1973), and this is at odds with the LFG desideratum of psychological plausibility (Kaplan & Bresnan 1982: 173–174). However, this objection only carries weight insofar as the rules are applied during on-line processing; if they only apply in the lexicon, their computational power is irrelevant, since lexical entries are stored in memory. The discovery that complex predicates necessitate an analysis whereby argument structures can be assembled in the syntax (Butt 1995, Alsina 1996; Section 4.2.4 below) challenges this solution, however. Another way to neutralise the processing objection is by formally implementing lexical rules in such a way as to make them tractable, such as by treating two lexical entries related by lexical rule as a single lexical entry containing disjunctive specifications (cf. fn. 10). This might result in quite a gap between theoretical LFG and computational implementations (which again runs counter to the Competence Hypothesis of Kaplan & Bresnan 1982), but it does at least avoid intractability.

Although none of these objections may be insurmountable, lexical rules have nevertheless fallen out of favour in LFG. Lexical Mapping Theory has offered a fruitful alternative that avoids the formal and conceptual issues of lexical rules, and also goes further, by providing an account of the initial linking of arguments and GFs. Lexical rules have not entirely disappeared, however, and are still sometimes invoked to capture certain generalisations over the lexicon – see e.g. Bresnan et al.'s (2016: 315–319) analysis of possessors and gerundives. However, such generalisations can also be captured by using TEMPLATES (Dalrymple et al. 2004, Asudeh et al. 2013, Belyaev forthcoming(a) [this volume]), providing the possibility of doing away with lexical rules altogether.

---

[9]Of course, we may not expect formalism to constrain theory in this way (cf. Pollard 1997), and in that case this objection is of less concern.

[10]Note that this is not an inherent property of lexical rules *per se*; as a reviewer notes, in XLE (the computational implementation of LFG – Kaplan & Newman 1997, Crouch et al. 2011), lexical rules are implemented as disjunctions of functional descriptions, thereby restoring monotonicity. This approach has also been taken in some theoretical work in LFG, starting with Butt et al. (1997) – see Section 6.2 below. HPSG takes a different approach to lexical rules again, treating them as unary-branching rules in the type hierarchy (see e.g. Davis & Koenig 2021: 155ff.).

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

4 Classical LMT

Lexical Mapping Theory (LMT) arose in part as a result of dissatisfaction with the shortcomings and unconstrained nature of lexical rules (Bresnan 1990). LMT therefore attempts to offer a more principled and constrained theory of both argument alternations and initial argument-GF mappings. Since the foundational work in LMT (Levin 1986, Bresnan & Kanerva 1989, Bresnan & Zaenen 1990), the theory has undergone many alterations and extensions; some of these build on one other, some offer competing perspectives, and some are simply different ways of saying the same thing. In addition, some are mere extensions or minor tweaks, while others involve rebuilding the theory from the ground up. We feel it would be both convoluted and unilluminating to trace every divergent strand of research in the LMT tradition, and so in this section we try to present a single coherent version of the theory, which we call Classical LMT. In order to maintain this coherence, we will adapt and update analyses where necessary, provided this does not detract from the main goals of the work in question.

Classical LMT represents what many take to be the "canonical" version of mapping theory in LFG, and is the variety which often appears in textbook presentations of the framework (as in e.g. Dalrymple 2001: 202ff. Falk 2001b: ch. 4, Bresnan 2001: ch. 14, Bresnan et al. 2016: ch. 14, and Börjars et al. 2019: ch. 8; see also Butt 2006: pp. 117ff.). However, it has long since been recognised that the name "*Lexical* Mapping Theory" is inappropriate, since "the theory cannot apply exclusively to individual words" (Dalrymple 2001: 212): for example, complex predicates which are formed analytically nonetheless contribute a single (complex) argument structure, despite the fact they contain multiple lexemes (Mohanan 1994, Butt 1995, Alsina 1996; Section 4.2.4 below). For this reason, alternative names have been proposed for the theory, including Mapping Theory *tout court* (as in e.g. Kibort & Maling 2015), Functional Mapping Theory (Alsina 1996), and Linking Theory (Butt et al. 1997). We use "Classical LMT" as a cover term, and for consistency with the large body of literature that uses the moniker "LMT", but we do not thereby intend to deny the importance of the work on complex predicates which shows that LMT cannot apply exclusively in the lexicon.

Our presentation of Classical LMT in this section has two parts: in Section 4.1, we present the basic formal tools and theoretical assumptions which characterise Classical LMT, while in Section 4.2 we discuss several case studies which illustrate the application of the theory to some empirical challenges, some of which necessitate (minor) changes to the theory.

Table 1: Feature decomposition of grammatical functions

	$-r$	$+r$
$-o$	SUBJ	OBL$_\theta$
$+o$	OBJ	OBJ$_\theta$

4.1 The framework

In this section, we present the theoretical and formal tools which are used in Classical LMT. We begin in Section 4.1.1 by introducing the idea of decomposing grammatical functions by means of binary features, which underpins the LMT approach to mapping. In Section 4.1.2, we address the question of the initial (un-marked) mapping of arguments to GFs, something that was ignored in the lexical rule approach. Lastly, Section 4.1.3 discusses the Classical LMT approach to argument alternations.

4.1.1 Feature decomposition

In the theoretical world described above in Section 3, arguments are associated with GFs in the lexicon. If those arguments are realised by different GFs as the result of some alternation, like the passive, the original assignments have to be overwritten. As discussed, this means that argument alternations involve non-monotonic re-writing rules. The key innovation of Classical LMT allowing it to avoid this unhappy conclusion is to underspecify the mappings between arguments and GFs, by grouping GFs into natural classes. Each argument can then be associated with one of these natural classes, rather than a specific GF, thereby constraining but not totally determining its ultimate realisation. And since the groupings of GFs are supposed to be natural, this also answers the complaint of unconstrainedness levelled at the lexical rule approach: no longer can we replace a GF with any other; instead, the choice of GFs available to an argument is limited to a natural class.

To achieve this cross-classification, Classical LMT decomposes the GFs using two binary-valued features, $[\pm r]$ and $[\pm o]$ (Bresnan & Kanerva 1989: 24–25). The first, $[\pm r]$, refers to whether a GF is thematically restricted or not: OBJ$_\theta$ and OBL$_\theta$ are; SUBJ and OBJ are not. The second, $[\pm o]$, refers to whether a GF is objective or not: OBJ and OBJ$_\theta$ are; SUBJ and OBL$_\theta$ are not. This is illustrated in Table 1. Grammatical functions can now be described in terms of two features: SUBJ is $[-r, -o]$, OBJ is $[-r, +o]$, OBL$_\theta$ is $[+r, -o]$, and OBJ$_\theta$ is $[+r, +o]$.[11] Each individual feature can

---

[11]If we take this feature decomposition literally, then grammatical functions are no longer prim-

also be used to describe a pair of GFs, as seen in each of the two rows and two columns of Table 1. This is what enables the association of an argument with a limited natural class of GFs: in Classical LMT, arguments are linked to a single feature (by means to be explored in the next section), and thereby made compatible with two GFs. This is more permissive than the original LFG approach, where an argument is linked to a specific GF, but still limited: argument alternations can only map the argument to the *other* GF, not to any arbitrarily different GF.

Bresnan & Kanerva (1989: 25) claim that the pairings induced by the feature decomposition just described are natural classes. This is a large part of the explanatory appeal of Classical LMT, so it is worth dwelling on momentarily. In fact, this is an area where Classical LMT has received some criticism. Alsina (1996), for example, observes that the standard feature decomposition fails in both directions: it describes an unnatural class and also fails to capture an important natural one. The pair of GFs described by $[+r]$, namely OBJ_θ and OBL_θ, does not seem to form a natural class, in that there are no instances where arguments alternate between them. At the same time, the division between terms/direct GFs and nonterms/obliques has a number of linguistic reflexes (Dalrymple et al. 2019: 15–17), yet no single feature can pick out the terms, i.e. SUBJ, OBJ, and OBJ_θ, or the nonterms, i.e. OBL_θ (Alsina 1996: 29, fn. 9). For this reason, Alsina (1996: 19–20) suggests a different decomposition, according to the features $[\pm\text{subj}]$ and $[\pm\text{obl}]$.

On a related note, Findlay (2020: 130) and Asudeh (2021: 32) object to the "suspiciously circular" (*ibid.*) definition of $[\pm o]$. While it might be relatively clear what independent content $[\pm r]$ could have (being semantically restricted makes sense outside of the context of grammatical functions), it is much less clear what independent content $[\pm o]$ could possess: it identifies a GF as belonging or not to the set {OBJ, OBJ_θ}, but by virtue of no other property than membership of that set.

Despite these qualms, the cleavages induced by the $[r]$ and $[o]$ features remain in common usage, even if their interpretation is reimagined (e.g. Kibort 2014: 266 views $[+o]$ as picking out the complements from the non-complements, and $[-r]$ as picking out the core arguments from the non-core – see Section 5). The most significant reason for this is ultimately their success: the cross-classification in Table 1 has proved incredibly useful in describing a variety of argument structure

---

itives in the theory; instead, the features are. Butt (1995: 31) makes this claim explicitly. However, it is also possible to avoid this conclusion, and retain the primitive status of grammatical functions in LFG, by viewing such feature decomposition as merely descriptive, so that it cross-classifies the GFs but does not formally break them down (Butt et al. 1997, Findlay 2016: 298ff.; see Section 6.2 below).

phenomena in a diverse selection of languages – we will see some examples of this later in this section and especially in Section 4.2.

One potential immediate issue is that using two binary-valued features enable us to describe a four-way classification, but LFG's inventory of grammatical functions has more than four members. Of course, it is no problem that we omit ADJ and XADJ from consideration, since adjuncts are not involved directly in argument structure and mapping, being unable (by definition) to be selected by a predicate.[12] However, the two clausal GFs COMP and XCOMP, both argument GFs, are also missing from Table 1. In fact, and despite some countervailing voices (Dalrymple & Lødrup 2000, Lødrup 2012), many researchers have advocated for eliminating these GFs by assimilating them to one or more of the other complement GFs, *viz.* OBJ, OBJ$_\theta$, and OBL$_\theta$ (Zaenen & Engdahl 1994: 197–198, Alsina 1996, Alsina et al. 2005, Forst 2006, Berman 2007, Patejuk & Przepiórkowski 2016, Szűcs 2018). In that case, the omission of COMP and XCOMP from Table 1 is not a problem. Even if the clausal GFs are not eliminated entirely, it seems possible that the distinction between them and the other complement GFs could still be neutralised at the level of specificity required of mapping theory. We can therefore continue to assume that the four GFs in Table 1 are the only ones relevant for mapping.

Besides dividing up the GFs, the [r] and [o] features can also be used to order them. Bresnan & Zaenen (1990: 49) claim that the features indicate markedness of GFs, so that those which possess more negative-valued features are less marked than those which possess more positive-valued ones. This leads to the partial ordering known as the MARKEDNESS HIERARCHY:

(30) *The Markedness Hierarchy:*
     SUBJ > OBJ, OBL$_\theta$ > OBJ$_\theta$

SUBJ, bearing a negative value for both features, is the least marked GF, at the top of the hierarchy; OBJ$_\theta$, with two positive values, is the most highly marked, at

---

[12]In fact, it has been argued that there are such things as "obligatory adjuncts", given the existence of contrasts like the following, where the omission of the parenthetical material leads to ungrammaticality on the intended reading of the verb:

(i)   a. Cat behaves *(badly).
       b. Lister lives *(in space).
       c. This book reads *(well).

See Przepiórkowski (2016: 262–263) and references therein for further discussion and exemplification.

the bottom. Since OBJ and OBL$_\theta$ both have one negative- and one positive-valued feature, they sit in the middle, and are not ordered with respect to one another.[13] This hierarchy of GFs is important for the principles which Classical LMT uses to determine the ultimate mapping of arguments to GFs, to which we now turn.

4.1.2 Initial classification and mapping of arguments

Just as in Section 3, we assume that predicates are equipped with an argument structure that lists their syntactically-realisable arguments. (31) shows a simple example for *kick*:

(31)   *kick* ⟨   Agent   Patient   ⟩

Although Classical LMT still offers no formal consensus on its status or position in the architecture of LFG, such a list now starts to be referred to as A-STRUCTURE, as if it were a separate level of the parallel projection architecture (see Falk 2001b: 97–106 for some discussion). Arguments within a-structure are ordered according to their thematic role, following the thematic hierarchy introduced in (10), and repeated in (32) (Bresnan & Kanerva 1989: 23):

(32)   *The Thematic Hierarchy:*
        Agent > Beneficiary > Recipient/Experiencer
                                   > Instrument > Theme/Patient > Location

The most important function of this ranking in Classical LMT is simply to identify the most highly ranked argument, which we refer to as $\hat{\theta}$.[14] This is because of the observation that the most "prominent" thematic role often aligns with the most "prominent" GF, i.e. SUBJ (Fillmore 1968, Grimshaw 1990, Speas 1990). We will see how this is cashed out in Classical LMT below.

---

[13]Note that the order of GFs in this hierarchy differs from the typologically-motivated Functional Hierarchy, which Dalrymple et al. (2019: 11) present as the standard in LFG (based on the Accessibility Hierarchy of Keenan & Comrie 1977):

   (i)   *The Functional Hierarchy:*
          SUBJ > OBJ > OBJ$_\theta$ (> XCOMP, COMP) > OBL$_\theta$ (> XADJ, ADJ)

Notably, OBJ$_\theta$ outranks OBL$_\theta$ in (i), while the opposite is true in (30).

[14]This is also sometimes called the THEMATIC SUBJECT or A-STRUCTURE SUBJECT, and has also been equated with the concept of LOGICAL SUBJECT. Such a notion of "most thematically prominent argument" has been shown to play a role outside of mapping theory as well, such as in determining the antecedent of a reflexive (Dalrymple 1993, Joshi 1993, Mohanan 1994, Manning & Sag 1999).

Whereas arguments were previously associated with a specific GF in the lexicon, in Classical LMT they are associated with a single [±o/r] feature instead (i.e. with a *pair* of GFs). In early versions of LMT, such as Bresnan & Kanerva (1989: 25–26) or Bresnan & Moshi (1990), this is achieved by intrinsic connections between specific named thematic roles and features, as in (33), from Bresnan & Moshi (1990: 168):

(33) *Intrinsic classifications:*

Agent	Theme/Patient	Location
[−o]	[−r]	[−o]

This is based on typological observations about common realisations of various thematic roles across languages: cross-linguistically, for instance, Themes/Patients canonically alternate between the unrestricted GFs, i.e. subject and object, while other roles like Agent and Location canonically alternate between the non-object functions, i.e. subject and oblique (Bresnan & Kanerva 1989: 26). There is no principled limit on which roles might receive intrinsic classifications like this.

In subsequent work in Classical LMT, this open-endedness is rejected, and the initial classification principles are reduced to three (Bresnan & Zaenen 1990: 49; cf. also Her 2003, 2013; and see Bresnan et al. 2016: 331 for a contemporary textbook presentation), claimed to be general across languages:[15]

(34) *Intrinsic classifications (general):*

patientlike roles:	secondary patientlike roles:	other roles:
θ	θ	θ
[−r]	[+o]	[−o]

While this is an improvement in terms of theoretical parsimony, there is a cost in terms of explicitness. Asudeh (2021: 32), for instance, complains that the notion of being "patientlike" is "obscure", noting that "it's not clear what the conditions are for meeting the criterion of being 'like' a patient".

Let us assume, however, that it is clear enough when a role is patientlike or not. What of the secondary patientlike roles? Where verbs have more than one patientlike argument, as in ditransitives, one of the two may be "secondary" in the sense of Dryer (1986), and this argument will be marked as [+o]. Such languages are called ASYMMETRICAL OBJECT LANGUAGES, in contrast with SYMMETRICAL OBJECT LANGUAGES, which permit multiple patientlike roles to be marked

---

[15]"θ" is used to stand for any thematic role, since these principles no longer refer to specific roles.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

[−r] (see Bresnan & Moshi 1990 and Section 4.2.3 below). Even within asymmetrical object languages, there is variation in which of the two arguments counts as primary or secondary – indeed, a single language can permit both possibilities (see discussion of English *give* below).

Given these basic assignments, the a-structure of our simple transitive verb *kick* will be as follows:

(35)   *kick*   ⟨   Agent   Patient   ⟩
                    [−o]   [−r]

There is one patientlike role, namely the Patient itself, so this is marked [−r]; the one other role is marked [−o], according to the third, "elsewhere" principle in (34).

To resolve these single features to fully-specified GFs, Classical LMT makes use of two MAPPING PRINCIPLES:[16]

(36)   *Mapping Principles*:

    a. Subject roles:

        i. $\hat{\theta}$ is mapped onto SUBJ when initial in the a-structure;
            [−o]

        otherwise:

        ii. θ is mapped onto SUBJ.
           [−r]

    b. Other roles are mapped onto the lowest featurally compatible function on the Markedness Hierarchy in (30).

As mentioned, the most thematically prominent argument, $\hat{\theta}$, is strongly associated with the SUBJ position; Mapping Principle (a-i) captures this, and requires that a non-patientlike $\hat{\theta}$ maps to SUBJ where possible. The constraint that $\hat{\theta}$ be leftmost in the a-structure is to account for the presence of non-thematic arguments which might take precedence in mapping to SUBJ. For example, the a-structure of a raising verb like *seem* is as shown in (37) (Zaenen & Engdahl 1994: 200):

(37)   *seem*   __   ⟨   Proposition   ⟩
                 [−r]   [−o]

---

[16]We follow the formulation of Bresnan et al. (2016: 334); for the first appearance of these principles, see Bresnan & Zaenen (1990: 51).

Although *seem* only takes a single semantic argument, the Proposition it embeds, this argument cannot surface as the subject, and the verb instead takes a non-thematic, expletive subject:[17]

(38) a. * That Kira smiled seemed.

      b. It seemed that Kira smiled.

For this reason, (37) contains two argument slots, although one is devoid of semantic content and is therefore marked as $[-r]$, since a non-thematic argument, by definition, cannot be semantically restricted. The highest thematic role, $\hat{\theta}$, is still the Proposition, and it is marked $[-o]$, but because it is no longer initial in the a-structure, it is not mapped to subj by Mapping Principle (a-i), leaving the expletive argument available to map to subj by Principle (a-ii).

In addition to the Mapping Principles in (36), there are two other well-formedness conditions on mapping, FUNCTION-ARGUMENT BIUNIQUENESS (Bresnan 1980: 112), and the SUBJECT CONDITION (Baker 2006 [1983], Bresnan & Kanerva 1989: 28):[18]

(39) *Function–Argument Biuniqueness:*
      Each a-structure role must be associated with a unique function, and vice versa.

(40) *The Subject Condition:*
      Every predicator must have a subject.

The first condition ensures that a predicate cannot select for multiple of the same GF, and that a single argument cannot be realised by multiple GFs of the same predicate.[19] The second represents a supposed language universal, that all predicates possess subjects – even when these are not overtly expressed. There have been some doubts about the universality of this claim (see e.g. Bresnan & Kanerva 1989: 28, fn. 37, Bresnan et al. 2016: 334, fn. 9, Kibort 2006, and references therein), so it may be more appropriate to see this as a parameter which varies by language.[20]

---

[17]Of course, there is also the "raised" alternative *Kira seemed to smile*. See Zaenen & Engdahl (1994) and Dalrymple et al. (2019: ch. 15) for the treatment of raising in LFG.

[18]Once again, we take the specific wording from Bresnan et al. (2016: 334).

[19]The first part of this is already barred by the f-structure well-formedness condition called Consistency (Dalrymple et al. 2019: 53–54), which follows from the functional nature of f-structure: each attribute at f-structure, such as a GF like subj or obj, can only have a single value.

[20]Kibort (2004: 358–359) reworks the Classical LMT Mapping Principles in such a way that she can do without the Subject Condition altogether – see Section 5 for more details.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Note that these well-formedness conditions are more important in early LMT work, such as Bresnan & Kanerva (1989), since this version of the theory does not include explicit Mapping Principles like (36). Instead, through a richer theory of intrinsic and default assignment of features to arguments, a number of mappings are made possible, which are then filtered down to the unique solution by Function-Argument Biuniqueness and the Subject Condition (Bresnan & Kanerva 1989: 28ff.). In the sense that this involves positing fewer rules, it is a simpler theory – but the rules it does include are more specific (i.e. referring to particular thematic roles by name), making it less general overall.

Let us return now to the example of a simple transitive predicate like *kick* and see how the Mapping Principles apply in practice. Since Agent outranks Patient on the Thematic Hierarchy, the Agent is identified as $\hat{\theta}$; since this argument is also initial in the a-structure, it is therefore mapped to subj. The remaining argument, the $[-r]$ Patient, then maps to the lowest compatible GF on the Markedness Hierarchy: the lowest $[-r]$ GF is obj. This correctly gives us the active voice mapping whereby the Agent is realised as the subject, and the Patient as the object:

(41)   *kick*   ⟨   Agent   Patient   ⟩

              $[-o]$ $[-r]$

               | |

            SUBJ   OBJ
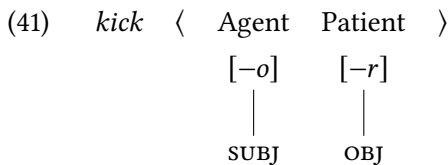
What of other predicate types?[21] Intransitives should have their single argument mapped to subj. The initial feature assignment to this argument will depend on whether the predicate is unaccusative or unergative (Perlmutter 1978):

---

[21]We consider only verbal predicates in this chapter. This footnote offers a selection of references for the reader interested in learning more about argument structure and mapping phenomena within the nominal domain. The most prominent idea, proposed by Rappaport (2006 [1983]), is that nominals derived from verbs inherit that verb's argument structure, but that the possibilities for mapping are more constrained within the noun phrase – for example, the functions subj and obj are not available to the dependents of nouns (cf. *Luke destroyed the Death Star* and *Luke's destruction of the Death Star*). This perspective remains the dominant one – see e.g. Laczkó (2000, 2003, 2007), Kelling (2003), Chisarik & Payne (2001, 2003) – but some have instead argued that nominals either don't have argument structures, or that, where they do, they can differ from the corresponding verbal ones (Ramchand 1997, Lowe 2017, Taylor 2023). Börjars & Lowe (forthcoming) [this volume] provide a useful contemporary summary of the issues.

A wide range of languages have been studied in LFG with respect to nominal argument structures and their mapping possibilities: see Saiki (1987) on Japanese, Markantonatou (1995) on Modern Greek, Laczkó (2000, 2003, 2004, 2010) on Hungarian, Falk (2001a) on Modern Hebrew, Kelling (2003) on French, Sulger (2013) on Hindi-Urdu, Lowe (2017) on Sanskrit and other early Indo-Aryan languages, and Taylor (2023) on Old English.

since the single argument of an unaccusative is patientlike, it will be assigned [−r]; unaccusatives, on the other hand, have more agentlike arguments, which will therefore be assigned [−o]. However, in both cases this will result in the correct mapping (in the simple, active case): for the unaccusative verb, (42), Mapping Principle (a-ii) applies, while for the unergative (43), Principle (a-i) does the job.

(42) *fall* ⟨ Patient ⟩
                 [−r]
                  |
                 SUBJ

(43) *run* ⟨ Agent ⟩
                 [−o]
                  |
                 SUBJ
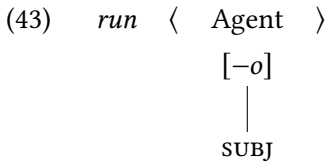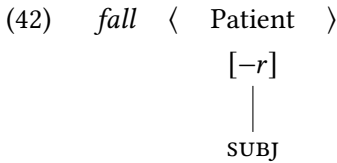
Ditransitives like *give* are slightly more complicated. They of course have three arguments in their a-structure:

(44) *give* ⟨ Agent   Beneficiary/Recipient   Theme ⟩
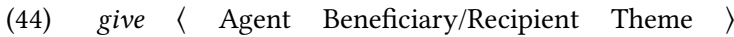
Following the usual initial classifications, the Theme, as a patientlike argument, is linked to [−r], and the Beneficiary/Recipient and Agent both receive the "elsewhere" [−o] feature. As per the Mapping Principles, the Agent, an a-structure-initial, [−o]-valued, $\hat{\theta}$ argument, is mapped to SUBJ. The Beneficiary/Recipient maps to the lowest [−o] GF, which is OBL$_\theta$, while the Theme maps to the lowest [−r] GF, OBJ. This gives us one correct mapping for *give*, illustrated in a sentence like *Peter gave a present to Harriet*.

(45) *give* ⟨ Agent   Beneficiary/Recipient   Theme ⟩
                 [−o]   [−o]   [−r]
                  | | |
                 SUBJ   OBL$_\theta$ OBJ
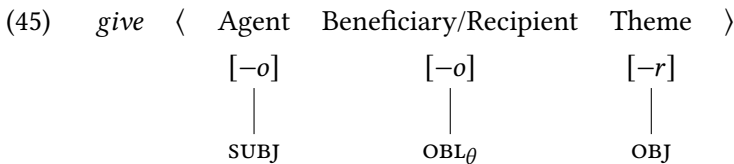
But of course there is another way of realising the arguments of a ditransitive like *give*: the dative-shifted version, illustrated in *Peter gave Harriet a present*. Since this involves the same thematic roles, this alternation cannot be derived in Classical LMT without some further stipulation (Kibort 2008: 314). It seems that we can choose to view the Beneficiary/Recipient as patientlike (cf. Toivonen 2013), in which case it is assigned [−r] by the intrinsic classification rules

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

(Bresnan 2003: 14–15; cf. also Bresnan et al. 2016: 337–340). Now, English is an asymmetrical object language, which means it does not permit the presence of two [−r] arguments at a-structure (see Section 4.2.3), and so the (lower-ranked) Theme must instead be marked [+o], as a secondary patientlike argument, per (34). The Agent receives the "elsewhere" [−o] specification as usual, giving us the following a-structure and GF-mapping:

(46) *give* ⟨ Agent   Beneficiary/Recipient   Theme ⟩
           [−o]   [−r]   [+o]
            | | |
          SUBJ   OBJ   OBJ$_\theta$

This is the double-object version of *give*: the Agent is mapped to SUBJ as usual, then the other arguments are mapped to the lowest compatible GFs, in this case OBJ for the Beneficiary/Recipient (the lowest [−r] GF) and OBJ$_\theta$ for the Theme (the lowest [+o] GF). So, Classical LMT can account for the dative shift alternation, but only with the initial stipulation that the Beneficiary/Recipient can be viewed as patientlike, and hence assigned [−r] at a-structure. Indeed, morphosemantic alternations in general are problematic for Classical LMT, a shortcoming which Kibort (2007, 2014) attempts to rectify, and which we will examine in more detail in Section 5. For now, though, we consider the well-developed Classical LMT account of (morphosyntactic) alternations.

4.1.3 Argument alternations

Argument alternations in Classical LMT are handled by adding extra specifications to arguments – in this way information is only added, not removed, meaning that "the computational requirement of monotonicity can be met even in the domain of relation changes" (Bresnan 1990: 650).

One common mechanism is that of SUPPRESSION, illustrated schematically in (47):

(47) θ
        |
        ∅

This prevents an argument from being mapped to a GF at f-structure, and existentially quantifies over the argument in the semantics (though it does allow the possibility of the argument being realised by an adjunct like the English *by*-phrase

which can express the Agent of a passive, so the quantification only applies if the argument remains unexpressed). Suppression is restricted to unmarked arguments, i.e. those pre-specified with a negatively-valued feature at a-structure (Alsina 1999; see also Bresnan et al. 2016: 338–340 for a relevant example), a principle known as RECOVERABILITY OF SUPPRESION (Bresnan et al. 2016: 333).

Lexical rules involving deletion can be recast in terms of suppression. For example, instead of deleting an OBJ, as in (25), intransitivisation involves suppression of a Theme/Patient argument:

(48)   Intransitivisation: Theme/Patient
$$|$$
$$\emptyset$$

And rather than deleting or re-writing the SUBJ and changing an OBJ to a SUBJ, as in (27), passivisation simply involves a single process, *viz.* the suppression of the highest thematic role:

(49)   Passivisation: $\hat{\theta}$
$$|$$
$$\emptyset$$

This simplified analysis of passivisation works because of the general system of mapping assumed in Classical LMT. In a standard two-place predicate like *kick*, the highest, Agent argument will be $[-o]$, while the next, Patient argument will be $[-r]$. If the Agent argument is suppressed, Mapping Principle (a-i) will not apply, and instead Principle (a-ii), which maps a $[-r]$ argument to SUBJ, will step in, correctly promoting the Patient argument, without any need for further stipulation:[22]

---

[22]The way the Mapping Principles are written, it seems to us that argument suppression should lead to a contradiction. Assuming the Principles are intended to be declarative rather than procedural, then (50) would seem to violate Mapping Principle (a-i), since it is not true that a $[-o]$, a-structure-initial $\hat{\theta}$ is mapped onto SUBJ: instead, it is not mapped to anything; and the same goes for intransitivisation: the suppressed Theme/Patient argument in a transitive will not be mapped to the lowest featurally compatible function on the Markedness Hierarchy, contrary to Principle (b). Perhaps suppression removes an argument from consideration at a-structure altogether, but in that case it would not be monotonic. One solution would simply be to add the rider "unless suppressed" to each of the Mapping Principles, but this seems far from parsimonious.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

(50)   *kicked$_{PASSIVE}$*   ⟨   Agent   Patient   ⟩

           [−o]   [−r]

             | |

             ∅   SUBJ

Passivisation also correctly applies to ditransitives in both their a-structure realisations. For example, suppressing the Agent in the non-shifted version, repeated in (51), results in the Theme being promoted to SUBJ, by Mapping Principle (a-ii), since it is a [−r] argument.

(51)   *give*   ⟨   Agent   Beneficiary/Recipient   Theme   ⟩

               [−o]   [−o]   [−r]

                 | | |

                 ∅   OBL$_θ$ SUBJ

This gives us the correct alternation, illustrated in (52), where the Beneficiary/Recipient remains an OBL$_θ$ (since this is still the most marked [−o] GF):

(52)   a.   Peter gave a present to Harriet.
       b.   A present was given to Harriet (by Peter).

On the other hand, when the Agent is suppressed in the dative-shifted version, the Beneficiary/Recipient is promoted instead, since it is now the [−r] argument, while the Theme remains an OBJ$_θ$ (since this is still the most marked [+o] GF):

(53)   *give*   ⟨   Agent   Beneficiary/Recipient   Theme   ⟩

               [−o]   [−r]   [+o]

                 | | |

                 ∅   SUBJ   OBJ$_θ$

This again accords with the facts:[23]

(54)   Peter gave Harriet a present.

(55)   Harriet was given a present (by Peter).

---

[23]For those dialects where *%A present was given Harriet (by Peter)* is grammatical, something more needs to be said, of course. It is possible the Asymmetrical Object Parameter (Bresnan & Moshi 1990) is not in force in these varieties of English (see Section 4.2.3 for more on the AOP).

92

Notice that because Mapping Principle (b) requires that an argument be mapped to the *lowest* compatible GF on the hierarchy, the [+o] argument of such double object verbs remains an OBJ_θ in the passive, and is not, for example, "promoted" to OBJ. That this is the correct result is not at all obvious from English data alone: the usual test for OBJ-hood is the possibility of promotion through passivisation, but we cannot passivise a passive. In the absence of any morphological marking of the distinction between OBJ and OBJ_θ, there is no obvious way to tell which of these two GFs *a present* bears in example (55).

Data from other languages, however, such as the Bantu language Chicheŵa, support the Classical LMT analysis. Ditransitive verbs can be formed in Chicheŵa by applicativisation, and when the applied argument is a Beneficiary, it is assigned a [−r] classification at a-structure, while the Theme is assigned [+o], exactly as in the English double object construction, and resulting in the same GF assignments as we saw above (Alsina & Mchombo 1993: 28). In such Chicheŵa applicatives, only the OBJ (the Beneficiary) can be indexed by an object marker on the verb, while the OBJ_θ (the Theme) cannot (Bresnan & Moshi 1990; Alsina & Mchombo 1993: 22):[24]

(56) a. Chi-tsîru chi-na-wá-gúl-ir-á   m-phâtso (a-tsíkāna).
          7-fool   7S-PST-2O-buy-APPL-FV 9-gift   2-girls
          'The fool bought a gift for them (the girls).'

      b. * Chi-tsîru chi-na-í-gúl-ir-á   a-tsíkāna (m-phâtso).
          7-fool   7S-PST-9O-buy-APPL-FV 2-girls   9-gift

Now, given the a-structure assignments, we also observe the same passivisation pattern for Chicheŵa applicatives as for the English double object construction, with the Beneficiary OBJ being promoted to SUBJ (Alsina & Mchombo 1993: 29):

(57) Atsíkāna a-na-phík-ír-idw-á   nyêmba.
      2-girls   2S-PST-cook-APPL-PASS-FV 10-beans
      'The girls were cooked beans.'

Crucially, we now have a diagnostic to identify the GF of the remaining Theme argument: if it is promoted to OBJ, it should be compatible with the presence of an agreeing object marker on the verb; if it remains an OBJ_θ, then the use of the object marker will not be possible. In fact, use of the object marker in this construction is ungrammatical (Alsina & Mchombo 1993: 30):

---

[24]Object NPs indexed on the verb can be omitted, indicated here by parentheses. Numbers signify noun classes; S = subject marker; O = object marker; FV = final vowel.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

(58) * Atsíkāna a-na-zí-phík-ír-idw-á   (nyêmba).
        2-girls   2S-PST-10O-cook-APPL-PASS-FV 10-beans
        'The girls were cooked beans.'

This incompatibility shows that the Beneficiary argument here must still be an
OBJ$_\theta$, not an OBJ, and this therefore motivates Mapping Principle (b), where ar-
guments are linked to the *most* marked compatible GF (though the empirical
landscape may not be quite so straightforward as this single data point would
suggest: see Kibort 2008 for some discussion of the complexities).

Along with suppression, argument alternations can involve adding new argu-
ments to an a-structure, as in the Bantu applicative (Bresnan & Moshi 1990), or
the English benefactive (Toivonen 2013). For example, Toivonen (2013: 514) gives
the rule in (60) for the benefactive in English, which takes a transitive verb into
a ditransitive, as in (59):

(59) a. I'll pack some sandwiches.
      b. I'll pack the children some sandwiches.

(60) English benefactive: $\langle$ $\hat{\theta}$ Beneficiary/Recipient   Theme $\rangle$
                                      $[-o]$ $[-r]$ $[+o]$

with $\varnothing$ pointing down to Beneficiary/Recipient.

Note that the symbol $\varnothing$ is used differently here from above, where it represented
argument suppression. Here it captures the fact that the Beneficiary/Recipient is
added to an a-structure which otherwise contains only a Theme and a $\hat{\theta}$, whatever
role that may play; i.e. (60) adds the Beneficiary/Recipient where previously there
was no argument.

As well as adding or suppressing arguments, alternations can also involve con-
straining the mapping possibilities of arguments. This is what happens in loca-
tive inversion, for example. The relevant examples from Chicheŵa are repeated
in (61):

(61) a. Chi-tsîme chi-li ku-mu-dzi.
        7-well   7-be 17-3-village
        'The well is in the village.'

      b. Ku-mu-dzi ku-li chi-tsîme.
        17-3-village 17-be 7-well
        'In the village is a well.'

Bresnan & Kanerva (1989: 27) analyse the relevant process in the following terms:

(62) Locative inversion: ⟨ Theme … Location ⟩
                                          $[-r]$

That is, when a Location appears in the same a-structure as a Theme, assign it the specification $[-r]$ in addition to whatever its intrinsic feature assignment is. Let us see how this provides the contrast in (61).

In the relevant sense, the verb *-li* 'be' takes a Theme and a Location argument; as per the intrinsic specifications of (34), the patientlike Theme is assigned $[-r]$ and the other role is assigned $[-o]$. All things being equal, this will provide the mapping instantiated by (61a), where the Theme maps to SUBJ, by Mapping Principle (a-ii), and the Location maps to OBL$_\theta$, the lowest $[-o]$ GF.

(63) *-li* ⟨ Theme Location ⟩
                  $[-r]$ $[-o]$
                    | |
                  SUBJ   OBL$_\theta$

When we apply the additional assignment in (62), however, things change:

(64) *-li* ⟨ Theme Location ⟩
                  $[-r]$ $[-o]$
                    | $[-r]$
                    | |
                  OBJ   SUBJ

Here, the Location argument is fully specified as a SUBJ, meaning that the Theme is prevented from also being mapped to SUBJ, owing to Function-Argument Bi-uniqueness. Instead, it must map to the lowest available GF on the Markedness Hierarchy, namely OBJ. This gives us the mapping instantiated by (61b).

This section has served to provide a sampling of the different approaches to argument alternations in Classical LMT. By suppressing, adding, or further specifying arguments, the theory can give succinct accounts of a variety of different phenomena. To the extent that these simple descriptions make the correct predictions in conjunction with the underlying theory, this also serves as a vindication of the latter. Of course, we have hardly been able to do justice to such a rich literature in a handful of pages, but we hope to have illustrated the key technical points. In the following section, we provide a few more case studies, further showcasing areas where Classical LMT has provided elegant and illuminating analyses.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

4.2 Case studies and extensions

The framework of Classical LMT has been shown to offer an elegant solution to many thorny empirical issues, but it has also sometimes been necessary to expand or modify the theory in the face of empirical deficiencies or theoretical shortcomings. In this section, we discuss various topics which showcase the workings of Classical LMT.

4.2.1 Resultatives

As first observed by Simpson (2006 [1983]), resultative predicates in English can be applied to the objects of transitives or to the subjects of their corresponding passives, as shown in (65), and to the subjects of unaccusative intransitives but not of unergatives, as shown in (66) (examples from Bresnan & Zaenen 1990: 46):

(65)   a.   We pounded the metal flat.
         b.   The metal was pounded flat.

(66)   a.   The river froze solid.
         b.   * The dog barked hoarse.

The question then arises: how should we characterise all and only the arguments which can have resultatives predicated of them?

The generalisation cannot be based on surface grammatical function. For one thing, the data above show that both subjects and objects can take resultative predicates. What is more, only some subjects are implicated: (66b) is ungrammatical, and (65a) would be too if it were intended to mean that we pounded the metal until *we* were flat.

Given the contrast between unaccusative and unergative predicates, we might think instead to appeal to the thematic role of the arguments in question: perhaps resultatives can be applied to Themes, and not to Agents? This would account for the data in (65–66), but unfortunately there are other data which invalidate such a generalisation. Resultatives can also be applied to non-thematic arguments such as "fake reflexives", illustrated in (67), or "non-subcategorised objects" which do not stand in a direct semantic relation to the main verb, illustrated in (68) (examples from Bresnan & Zaenen 1990: 47):

(67)   a.   The dog barked itself hoarse.
         b.   We ran ourselves ragged.

(68)   a.   The dog barked us awake.

b. We ran the soles right off our shoes.

The tools of Classical LMT offer a straightforward solution to this descriptive challenge: the arguments in question are simply those which are assigned [−r] as their initial feature value at a-structure. For the Themes in (65–66), this follows from their being patientlike, while for the problematic arguments in (67–68) this follows from their being non-thematic (and so by definition semantically unrestricted). The more agentive subjects of transitive and unergative verbs will instead by classified as [−o] by the "elsewhere" condition, which sets them apart.

4.2.2 Proto-roles and unaccusativity

Another area where intrinsic classification of argument positions at a-structure has proved a more useful discriminator than other notions is in Zaenen's (1993) analysis of unaccusativity in Dutch. Before we consider the data, however, we first introduce Zaenen's innovative approach to intrinsic feature specification.

Rather than having to decide impressionistically whether an argument is "patientlike" or not, in order to decide whether it should be assigned [−r] or [−o] as its initial feature specification at a-structure, Zaenen (1993: 146–154) proposes to operationalise Dowty's (1991) notion of semantic PROTO-ROLE.

Dowty (1991: 571–575) envisages semantic roles as prototypes: arguments can possess a number of both proto-agent and proto-patient properties, with their behaviour depending on the balance between the two groups. This allows a fuzzier notion of semantic role, and avoids some of the definitional challenges of using named roles. Proto-agentivity and proto-patientivity are determined by a number of lexical entailments, including volition, change of state, and movement, which describe aspects of the relationship between participant and event (Dowty 1991: 572):

(69) *Proto-agent entailments:*

- volitional involvement in the event or state
- sentience (and/or perception)
- causing an event or change of state in another participant
- movement (relative to the position of another participant)
- exists independently of the event named by the verb

(70) *Proto-patient entailments:*

- undergoes change of state

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

- incremental theme[25]
- causally affected by another participant
- stationary relative to movement of another participant
- does not exist independently of the event, or not at all

Dowty (1991: 576) uses these proto-properties to determine the assignment of the subject and object GFs to arguments (the argument with more proto-agent properties becomes the subject, while the argument with more proto-agent properties becomes the object), but Zaenen (1993: 149) instead uses them to determine the intrinsic feature specification of an argument at a-structure: those that have more proto-agent properties will be classified as $[-o]$, while those that have more proto-patient properties will be classified as $[-r]$. This therefore captures the same general intuition as the Classical LMT intrinsic assignment principles in (34), namely that patientlike arguments are $[-r]$ and others are $[-o]$, but does so in a way which makes it more explicit what criteria an argument has to satisfy to count as patientlike. (Of course, determining whether an argument satisfies the proto-properties can also sometimes be rather impressionistic, but many are clear-cut enough to at least afford one an analytical toehold.)

A problem arises when an argument possesses an equal number of proto-agent and proto-patient properties (including zero). Dowty (1991: 576) proposes that in this situation both mappings are available. Zaenen (1993: 150) instead assumes that in such a case the argument is assigned $[-r]$. This is somewhat self-serving in that it gives her the correct results for Dutch (see below), but, as she observes, it does not seem unreasonable that it is precisely in areas such as this, where the distinctions are less clear cut, that languages vary, and so perhaps a degree of arbitrariness is unavoidable.

Let us now turn to the Dutch data which Zaenen (1993) uses these tools to analyse. Intransitive verbs in Dutch take different auxiliaries in the compound past tense depending on whether they are unaccusative or unergative. The unergatives take *hebben* 'have' and the unaccusatives take *zijn* 'be':

(71) **Unergative verbs:**

    a. Hij heeft/*is gelopen.
        he has/is   run
        'He has run.'

---

[25]Dowty (1991: 588) defines an incremental theme as "an NP that can determine the aspect of the sentence [...]; the event is 'complete' only if all parts of the NP referent are affected (or effected)". For example, in *Chrisjen ate a pistachio*, the eating event is only complete once all (edible) parts of the pistachio are eaten.

    b. Ze heeft/*is getelefoneerd.
       she has/is   telephoned
       'She has telephoned.'

(72) **Unaccusative verbs:**
    a. Ze is/*heeft overleden.
       she is/has   died
       'She has died.'

    b. Hij is/*heeft gevallen.
       he is/has   fallen
       'He has fallen.'

This also correlates with another contrast: the possibility of using the past participle as a pre-nominal modifier. This is impossible with the unergative, *hebben*-taking verbs, but perfectly productive with the unaccusative, *zijn*-taking verbs:

(73)   a. * de gelopen/getelefoneerd man
          the run/telephoned   man

      b. de overleden/gevallen vrouw
          the deceased/fallen   woman
          'the deceased/fallen woman'

    Now, if the intransitives were the only verbs we had to consider here, then a semantic explanation would be possible. For one thing, the single argument of an unaccusative is generally Theme/Patient-like. Zaenen (1993: 132–136) also discusses other semantic criteria which distinguish the two classes of verbs. However, a class of transitive verbs (those with an experiencer argument) also exhibit the same syntactic split, despite having different semantics. Firstly, some take *hebben* and some take *zijn* in the compound past tense:

(74)   a. Dat is/*heeft me jarenlang goed bevallen.
          that is/has   me for.years well pleased
          'That has pleased me well for years.'

      b. Hij heeft/*is me jarenlang geïrriteerd.
          he has/is   me for.years irritated
          'He has irritated me for years.'

And this distinction once again maps onto a difference in the use of the past participle as a pre-nominal modifier. When the past participles of those verbs

that take *zijn* are used pre-nominally, their head noun can be understood as the equivalent of their active voice subject, whereas this is not the case for those that take *hebben*:

(75) het hem goed bevallen boek
      the him well pleased book
      'the book that pleased him well'

(76) a. de geïrriteerde jongen
          the irritated   boy
          'the irritated boy'
      b. # de geïrriteerde fouten
          the irritated   mistakes
          'the mistakes that were irritated', not 'the mistakes that caused irritation'

But here the semantic explanation is not available: the subject of a verb like *bevallen* 'please/suit' is not a Theme/Patient, but rather a Stimulus or equivalent. And Zaenen (1993: 144) notes that "if there are any semantic properties that distinguish the two classes of experiencer verbs under consideration, they are not the same as the ones distinguishing the two classes of intransitives".

In fact, once again the solution is to look at intrinsic assignment of features at a-structure. The subjects of verbs like *bevallen* do not, in Zaenen's (1993: 149) view, possess any proto-agent or proto-patient entailments; in the event of a tie, Zaenen (1993: 150) assumes that the argument is assigned $[-r]$, and so these arguments are treated as being patientlike. We now have an explanation for the shared unaccusative/unergative split across intransitives and transitives. Just as with resultatives, the presence of an $[-r]$ argument is the significant factor: verbs in which the intrinsically $[-r]$-marked argument becomes subject take the auxiliary *zijn* (otherwise verbs take *hebben*), and the head noun of the pre-nominal participle corresponds to the $[-r]$ argument – this makes such participial uses simply impossible for unergative intransitives, which have no $[-r]$ argument, and means that the head noun corresponds to the "logical object" of transitives.

4.2.3 Double object constructions

The world's languages are divided in how they treat ditransitive predicates. For some, both objects of a ditransitive are treated equally: for example, either can be promoted to subject by passivisation, flagged by object marking on the verb, etc.

As mentioned above, these languages are called symmetrical object languages. Other languages, called asymmetrical object languages, exhibit strong differences between "primary" and "secondary" objects, whereby only one object is eligible for promotion by passivisation, flagging by object-marking on the verb, etc. This distinction was first drawn as a result of work on the Bantu languages (e.g. Gary & Keenan 1977, Kisseberth & Abasheikh 1977, Baker 1988), where the divide is particularly clear: since these languages have a productive process of applicativisation, ditransitive predicates are very frequent, and a number of grammatical features are sensitive to objecthood.

To illustrate the contrast between symmetrical and asymmetrical object languages, we consider two languages from the Bantu family: Kichaga and Chicheŵa. Kichaga is a symmetrical object language, and so either of the post-verbal arguments in the active can be promoted to subject by passivisation:

(77)   a.   N-ắ- í-lyì-í-à   m̀-kà   k-élyà
            FOC-1S-PRS-eat-APPL-FV 1-wife 7-food
            'He is eating food for/on his wife.'

   b.   M̀-kà   n-ắ-í-lyì-í-ò   k-élyâ
        1-wife FOC-1S-PRS-eat-APPL-PASS 7-food
        'The wife is being eaten food for/on.'
        (i.e. 'The wife is being benefitted/adversely affected by someone eating food.')

   c.   K-élyà k-í-lyì-í-ò   m̀-kà
        7-food 7S-PRS-eat-APPL-PASS 1-wife
        'The food is being eaten for/on the wife.'

Chicheŵa, on the other hand, is an asymmetrical object language. Here, only the immediately post-verbal argument in the active can be promoted to subject in the passive (Baker 1988: 248):

(78)   a.   Kalulu a-na-gul-ir-a   mbidzi nsapato.
            hare   s-PST-buy-APPL-ASP zebras shoes
            'The hare bought shoes for the zebras.'

   b.   Mbidzi zi-na-gul-ir-idw-a   nsapato (ndi kalulu).
        zebras s-PST-buy-APPL-PASS-ASP shoes   by hare
        'The zebras were bought shoes (by the hare).'

   c.   * Nsapato zi-na-gul-ir-idw-a   mbidzi (ndi kalulu).
        shoes   s-PST-buy-APPL-PASS-ASP zebras   by hare
        'Shoes were bought for the zebras (by the hare).'

There are a number of other properties which correlate with the passivisation facts (Bresnan & Moshi 1990: 150–153). Either or both post-verbal arguments in Kichaga can be omitted if they are encoded on the verb by an object marker, for instance, while in Chicheŵa, only the immediately post-verbal Beneficiary argument can be encoded/omitted this way; Kichaga allows unspecified object deletion of the Patient in a ditransitive where Chicheŵa does not; Kichaga allows the Patient argument to be eliminated by reciprocal marking on the verb in the presence of any applied object, while this is not the case in Chicheŵa; and all of these properties can interact in different ways.

These patterns receive an elegant explanation in Classical LMT, by way of the Asymmetrical Object Parameter (AOP; Alsina & Mchombo 1990, Bresnan & Moshi 1990: 172). This is a well-formedness constraint on a-structures, parametrised so that some languages apply it (i.e. asymmetrical object languages) and others do not (i.e. symmetrical object languages).

(79) *Asymmetrical Object Parameter*

$$* \quad \theta \qquad \ldots \qquad \theta$$
$$\quad | \qquad\qquad\qquad |$$
$$\quad [-r] \qquad\qquad [-r]$$
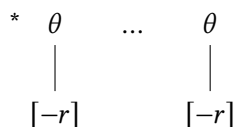
The AOP prohibits the presence of two intrinsically classified $[-r]$ arguments in the same a-structure: when it is in force, secondary patientlike arguments are assigned $[+o]$ by the intrinsic linking principles introduced in Section 4.1.2; when it is not, we permit multiple patientlike arguments to be assigned $[-r]$ instead. Let us consider how this can explain the passivisation facts shown in (77) and (78).

Chicheŵa is an asymmetrical object language, so the AOP is active. The a-structure for an applicative verb like we see in (78a) is therefore as follows:

(80) *gulira* 'buy-for' ⟨ Agent Beneficiary Theme ⟩
$$\qquad\qquad\qquad\qquad\quad [-o] \qquad [-r] \qquad [+o]$$
$$\qquad\qquad\qquad\qquad\quad | \qquad\qquad | \qquad\qquad |$$
$$\qquad\qquad\qquad\qquad\;\; \text{SUBJ} \qquad \text{OBJ} \qquad \text{OBJ}_\theta$$
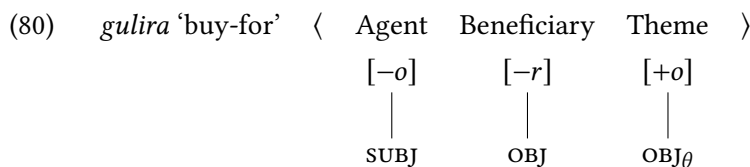
Just as with the English ditransitive above, we interpret the Beneficiary as patientlike, and so assign it the intrinsic feature $[-r]$. By the AOP, the second patientlike argument cannot also be marked $[-r]$, so it is instead classified as $[+o]$. This leads to the (correct) mapping shown in (80).

In the passive, only the Beneficiary is eligible for promotion to SUBJ when the Agent is suppressed, since the $[+o]$ Theme is featurally incompatible. This explains the contrast between (78b) and (78c).

(81) *guliridwa* 'buy-for_PASSIVE' ⟨ Agent Beneficiary Theme ⟩
                                      $[-o]$ $[-r]$ $[+o]$
                                        ∅   SUBJ   OBJ$_\theta$

Now consider Kichaga. Since it is a symmetrical object language, we are free to ignore the AOP ban on having two intrinsically $[-r]$-marked arguments. However, if we do, then we run into trouble in the active:

(82) *lyìíà* 'eat-for' ⟨ Agent Beneficiary Patient ⟩
                          $[-o]$ $[-r]$ $[-r]$
                          SUBJ   OBJ   *

Since the Agent will be mapped to SUBJ, we are left with only one remaining $[-r]$ GF to share between two arguments. So here Kichaga must take the same option as Chicheŵa of assigning the non-Beneficiary argument $[+o]$ instead:

(83) *lyìíà* 'eat-for' ⟨ Agent Beneficiary Patient ⟩
                          $[-o]$ $[-r]$ $[+o]$
                          SUBJ   OBJ   OBJ$_\theta$

However, in the passive, things are different. Now that the Agent is not mapped to any GF, there are still two $[-r]$ GFs available. This means the unrestricted intrinsic mapping of two arguments to $[-r]$ is possible, and will in fact lead to two possible final mappings:

(84) *lyìíò* 'eat-for_PASSIVE' ⟨ Agent Beneficiary Patient ⟩
                                $[-o]$ $[-r]$ $[-r]$
                                ∅   SUBJ/OBJ   OBJ/SUBJ

This is exactly the right prediction, since both (77b) and (77c) are grammatical.

The other properties can also be made to follow from the possibility of having multiple $[-r]$ arguments or not. Recall that the argument structure operation of suppression is limited to unmarked arguments (those that possess negatively-valued intrinsic features) – it then follows that e.g. unspecified object deletion applies more freely in symmetrical object languages, which can have more arguments with negatively-valued features than asymmetrical object languages.

4.2.4 Complex predicates

Complex predicates are predicates which syntactically head single clauses, but whose meanings incorporate multiple semantic heads and which therefore have complex argument structures. They have been at the centre of LFG work on argument structure and mapping theory since the earliest days, and have consistently drawn a great deal of attention in the literature (e.g. Ishikawa 1985, Alsina 1992, 1996, Butt 1995, 2014, Mohanan 1994, Matsumoto 1992, 1996, Andrews & Manning 1999, Lowe 2016, Lovestrand 2020, among many, many others; see also Dalrymple et al. 2019: 351–352 for an overview of the range of cross-linguistic work on complex predicates carried out in LFG).[26] As one might expect, therefore, this work has also led to various innovations and extensions of Classical LMT. In this section, we discuss two of these: the idea that one a-structure can be embedded inside another, with appropriate fusion of overlapping arguments, and the claim that this a-structure composition can take place in the syntax proper, not just in the lexicon, thus putting paid to the "lexical" aspect of Lexical Mapping Theory.

The first of these points can be seen by considering causatives in Chicheŵa (Alsina 1992). Verbs containing the causative suffix *-íts* add an additional Causer argument which, in the active, surfaces as the subject, with the previous subject being demoted, either to object or oblique status (Alsina 1992: 518):

(85) Nŭngu   i-na-phík-íts-a   kadzidzi maûngu.
     9.porcupine 9s-PST-cook-CAUS-FV 1a.owl   6.pumpkins
     'The porcupine made the owl cook the pumpkins.'

---

[26]There has also been extensive work on computational grammars for LFG that can handle complex predicates, with a particular focus on Hindi-Urdu: see Butt et al. (2003, 2012), Butt & King (2007), Bögel et al. (2009), Sulger (2013).

Another strand of research worth highlighting studies the consequences of complex predicates for the syntax-semantics interface: see Dalrymple, Hinrichs, et al. (1993), Kaplan & Wedekind (1993), Andrews & Manning (1999), Andrews (2007), Homola & Coler (2013), Lowe (2015).

(86) Nŭngu   i-na-phík-íts-a   maûngu   (kwá kádzīdzi).
      9.porcupine 9s-PST-cook-CAUS-FV 6.pumpkins to   1a.owl
      'The porcupine had the pumpkins cooked by the owl.'

Now, we might imagine that such causative forms have a simple a-structure, containing three argument positions for the Causer, Causee, and original Patient (here *maûngu*, 'pumpkins'). Instead, Alsina (1992: 521) suggests they have a complex argument structure, formed by embedding the base verb's a-structure into the a-structure of the CAUSE predicate, whose Patient is then merged with one of the arguments of the base predicate:

(87) CAUSE ⟨ Agent   Patient   PRED ⟨...θ...⟩ ⟩
                               ⎵⎵⎵⎵⎵⎵
                               caused event

Where the base predicate has more than one argument, this means there are multiple possibilities for this ARGUMENT FUSION: for instance, the causative's Patient argument may fuse with either the Agent or Patient of *phīka* 'cook'. Alsina (1992: 523–524) claims that this is precisely the difference between the two realisations in (85) and (86). In (85), the causative Patient is combined with the embedded verb's Agent, meaning the Causer's goal was to make the owl carry out the cooking; this sentence, but not (86), is therefore a possible answer to the question "What did the porcupine do to the owl?". In (86), however, the causative Patient is fused with the embedded verb's Patient, meaning the Causer merely intended for the pumpkins to get cooked, but did not especially care whether the owl did it; this sentence, but not (85), is therefore a possible answer to the question "What did the porcupine do to the pumpkins?".

  The fact that an argument of the base predicate is the Patient of the causative morpheme itself has a number of effects For instance, although the verb *phīka* 'cook' normally allows deletion of its object, in its causative form this is not possible, showing that in this respect the object behaves like an argument of the causative morpheme, rather than of the base predicate (Alsina 1992: 524–525):

(88) Kadzīdzi a-na-phík-a   (maûngu).
      1a.owl   1s-PST-cook-FV 6.pumpkins
      'The owl cooked (the pumpkins).'

(89) Nŭngu   i-na-phík-îts-a   *(maûngu)   (kwá kádzīdzi).
      9.porcupine 9s-PST-cook-CAUS-FV 6.pumpkins   to   1a.owl
      'The porcupine had the pumpkins/something cooked (by the owl).'

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

At the same time, the fused argument is also sensitive to its thematic role within the embedded predicate – for example, if it is an Agent in the base predicate it cannot be extracted (e.g. by relativisation), whereas if it is a Patient then it can (Alsina 1992: 529–530). This mixed behaviour motivates the idea that two argument positions are fused in the a-structure of these complex predicates.

The assumption of argument fusion also allows a straightforward Classical LMT account of the mapping possibilities open to causatives in Chicheŵa. The alternation between (85) and (86), for example, follows naturally if we assume that when two arguments fuse it is only the higher one which receives its intrinsic feature assignment:[27]

(90) a. *phikītsa* 'cause to cook'

$$\langle \quad \text{Agent} \quad \text{Patient} \quad \langle \quad \text{Agent} \quad \text{Patient} \quad \rangle \quad \rangle$$
$$[-o] \qquad [-r] \qquad\qquad\qquad [+o]$$
$$\text{SUBJ} \qquad \text{OBJ} \qquad\qquad\qquad \text{OBJ}_\theta$$

b. *phikītsa* 'cause to cook'

$$\langle \quad \text{Agent} \quad \text{Patient} \quad \langle \quad \text{Agent} \quad \text{Patient} \quad \rangle \quad \rangle$$
$$[-o] \qquad [-r] \qquad [-o]$$
$$\text{SUBJ} \qquad \text{OBJ} \qquad \text{OBL}_\theta$$

When the causative Patient is fused with the embedded Agent, the embedded Patient is marked [+o] as a secondary patientlike argument (the higher Patient taking priority owing to its ranking in the a-structure). When it is fused with the embedded Patient instead, the embedded Agent now receives a [−o] specification, but since the higher Agent is leftmost in the a-structure, it will map to SUBJ, leaving this lower Agent to map to OBL$_\theta$ instead.

We can also see why the causatives of intransitives do not exhibit this same alternation – their Causee can only surface as an OBJ, never as an OBL$_\theta$:

(91) Chatsalĭra a-ku-nám-íts-á   (* kwá) mwǎina.
      1.Chatsalira 1S-PRS-lie-CAUS-FV   to   1.child
      'Chatsalira is making the child tell lies.'

---

[27]We diverge somewhat from Alsina's (1992) proposal here – albeit only in detail and not in spirit – in order to harmonise with the approach to mapping we introduced earlier.

This follows naturally from the argument structure facts: since the embedded predicate only has a single argument, that will necessarily be the argument that fuses with the causative Patient, and so it is mapped to OBJ, not to OBL$_\theta$:

(92) *namĭtsa* 'cause to lie'

⟨ Agent   Patient   ⟨   Agent   ⟩   ⟩

[−o]   [−r]

SUBJ   OBJ
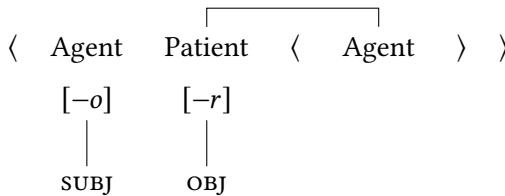
Chicheŵa forms causatives morphologically, and so the processes of a-structure composition and argument fusion can be thought of as taking place in the lexicon. However, some complex predicates are made up of multiple words, and so their argument structures must be built in the syntax rather than in the lexicon. Butt (1995), studying Hindi-Urdu permissive and aspectual constructions, and Alsina (1996), studying Romance causatives, were among the first to make this observation. We will illustrate the phenomenon with Hindi-Urdu data.

In Hindi-Urdu, complex predicates can be formed from a combination of a main verb and a light verb. In the case of so-called permissive complex predicates, the light verb in question is *de* 'let', homophonous with the lexical verb meaning 'give' (Butt 1995: 35). As with the causative morpheme, the light verb contributes its own arguments, which are added to and overlap with the arguments of the main predicate. For example, in (93), *saddaf=ko* is at once the "lettee" argument of the light verb *diyaa* and the "maker" argument of *banaane* 'make' (other arguments belong to only one verb: *anjum=ne* is only an argument of *diyaa* – she is the the one giving permission – and *haar* 'necklace' is only an argument of *banaane* – it is the thing being made).

(93) anjum=ne   saddaf=ko   haar   banaa-ne
      Anjum.F=ERG Saddaf.F=DAT necklace.M.NOM make-INF.OBL
      di-yaa.
      give-PERF.M.SG
      'Anjum let Saddaf make a necklace.'

The light verb and main predicate do not have to be adjacent or form a constituent at c-structure, so there is no sense in which they can be analysed as a single, morphologically complex word (Butt 1995: 46):

(94) a. anjum=ne saddaf=ko haar [**banaa-ne di-yaa**].

    b. anjum=ne **di-yaa** saddaf=ko [haar **banaa-ne**].

    c. anjum=ne [haar **banaa-ne**] saddaf=ko **di-yaa**.

Nevertheless, these sentences do not involve clausal embedding: with respect to agreement, anaphora, and control, they behave monoclausally (see Butt 1995: 36–43 for detailed evidence of this). That is, they have a flat f-structure, shown in (95):[28]

(95)
$$\begin{bmatrix} \text{PRED} & \text{`LET-MAKE}\langle\text{SUBJ,OBJ,OBJ}_{\text{GOAL}}\rangle\text{'} \\ \text{SUBJ} & [\text{PRED `ANJUM'}] \\ \text{OBJ}_{\text{GOAL}} & [\text{PRED `SADDAF'}] \\ \text{OBJ} & [\text{PRED `NECKLACE'}] \end{bmatrix}$$

This means the complex predicate must also have a single, composite a-structure:

(96)

de 'let/give' 〈   Agent   Goal   *banaa* 'make' 〈 Agent   Theme 〉 〉

              [−*o*]   [+*o*]   [−*r*]

              SUBJ OBJ$_{\text{GOAL}}$ OBJ

But this a-structure cannot be the property of any one word in the lexicon, since it combines information from two words, and the light verb can freely combine with various predicates. What is more, complex predicates can be recursively embedded – Butt et al. (2010) give an example involving four levels of embedding, for instance:

(97) taaraa-ne amu-ko (bacce-se) haathii   pinc kar-vaa
      Tara-ERG Amu-DAT child.OBL-INS elephant.M.SG.NOM pinch do-CAUS
      le-ne   dii-yaa.
      take-INF.OBL give-PRF.M.SG
      'Tara let Amu have the elephant pinched (by the child) (completely).'

The core meaning here is the noun-verb complex predicate made up of *pinc* 'pinch' and *kar* 'do'. This is then embedded under a causative predicate, which is hosted morphologically on this same light verb. Then we have a "completive" light verb *le* (with the lexical meaning 'take'). Finally, this whole complex is embedded under the permissive light verb *de*, which we saw above.

---

[28]The question of how the composite PRED value emerges here is an unanswered one – see Lowe (2016: sec. 2) for a sceptical review.

The conclusion such data must lead us to is that complex predicate formation is a productive, syntactic process, which means that we need to be able to combine a-structures on-line, outside of the lexicon. Apart from anything else, this means that the name "Lexical Mapping Theory" is a misnomer, since the theory must not apply only to individual words, but also to complex predicate-argument structures built up syntactically.

5 Kibort MT: incorporating morphosemantic alternations

We've now seen a sampling of the successes of and challenges for Classical LMT. In this section, we turn to a rather different view of LMT, that developed by Kibort over a series of papers (Kibort 2001, 2007, 2008, 2009, 2013, 2014, Kibort & Maling 2015), which purports to improve on Classical LMT in a number of respects, not least of which being its ability to handle morphosemantic alternations. We will refer to this theory as KIBORT MT.

As we saw in Section 4.1.2, the dative shift alternation poses a challenge for Classical LMT, in that the theory must assume two distinct initial assignments of features to arguments in order to be able to derive the two alternants. Other morphosemantic alternations are even more challenging. Consider again the *spray/load* alternation (Levin 1993: 50–51), illustrated in (98):

(98) a. Adam sprayed the paint on the wall.
      b. Adam sprayed the wall with the paint.

This is morphosemantic insofar as the entailments of the alternants differ: in each case, the participant corresponding to the OBJ is completely affected – i.e. in (98a) the paint is fully used up, while in (98b) the wall is totally covered. Once again, both alternants involve the same thematic roles, and so the basic Classical LMT a-structure will be the same for both:

(99)   *spray* ⟨ Agent Goal Theme ⟩

We would expect the Theme, being patientlike, to be assigned $[-r]$, and the other arguments to receive the default $[-o]$ assignment; this correctly produces the alternant in (98a), where the Theme surfaces as SUBJ, and the Goal as an OBL$_\theta$:

(100)   *spray* ⟨ Agent   Goal   Theme ⟩
                    $[-o]$ $[-o]$ $[-r]$
                      | | |
                    SUBJ   OBL$_\theta$ OBJ

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Producing the other alternant, in (98b), is much more difficult, however. Compared to (98a), we need the Goal and Theme to switch GFs: the former now surfaces as an OBJ, and the latter as an OBL$_\theta$. We could try the same trick as we did for ditransitive *give*, and say that the Goal argument counts as patientlike: this will allow us to classify it as $[-r]$, so that it can map to OBJ. But now the Theme will receive a $[+o]$ assignment as a secondary patientlike argument, which is incompatible with the $[-o]$ GF OBL$_\theta$. Indeed, patientlike arguments can only be classified as $[-r]$ or $[+o]$ by the intrinsic assignments in (34), which is precisely the opposite of what is needed to be compatible with the $[+r, -o]$ specification of OBL$_\theta$.

In Kibort's view, the problem arises because Classical LMT conflates syntactic arguments and semantic participants, representing both simultaneously in the list of arguments-*cum*-thematic roles. She proposes therefore to expand the domain of a-structure and mapping theory to include not only ARGUMENT-FUNCTION MAPPING, i.e. what we have been considering as the domain of mapping theory up to now, but also ARGUMENT-PARTICIPANT MAPPING.[29] This is illustrated in Figure 1, representing the typical active voice realisation of the Polish double object verb *dać* 'give' (cf. Kibort 2014: 265).[30]

semantic participants	Agent	Theme	Recipient	SEMANTIC VALENCY
	|	|	|	
arguments of the predicate	$\langle$ arg$_1$	arg$_2$	arg$_3$ $\rangle$	LEXICAL VALENCY
	|	|	|	
grammatical functions	SUBJ	OBJ	OBJ$_\theta$	FUNCTIONAL SUBCATEGORISATION
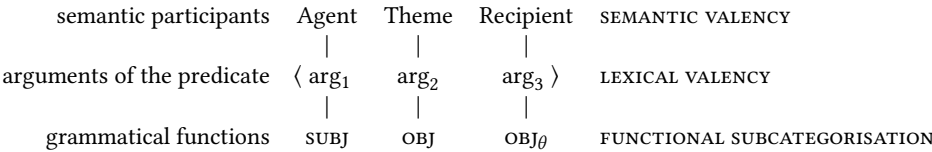
Figure 1: The separation of levels in Kibort MT

Before providing the Kibort MT solution to the *spray/load* puzzle, we first introduce the theory in more detail.

Kibort retains the Classical LMT mapping features $[\pm r]$ and $[\pm o]$, but, in keeping with the separation of syntax and semantics shown in Figure 1, she reinterprets them in purely syntactic terms, according to two traditional classifications of verbal dependents (Kibort 2014: 266):[31]

---

[29]In other works by Kibort, these are referred to as "argument-**to**-function/participant mapping", but since the connections are intended to be bidirectional, we omit the preposition here to minimise the procedural implications.

[30]As Kibort (2007: 252) points out, separating argument positions from semantic participants in fact goes back to early LFG work (such as Bresnan 1982), and has been argued for by others such as Grimshaw (1988: 1), Mohanan (1990), Ackerman (1991: 12, 1992: 57ff), Mohanan (1994: 15ff), Joshi (1993), Alsina (1996: 37), Falk (2001b: 105), and Ackerman & Moore (2013: 40ff).

[31]At least two other LFG linguists have proposed LMT feature sets which make no reference to semantic/thematic restrictions: Alsina (1996) and Hemmings (2012).

(101) $[-o]$ non-complements   (the "external" argument and oblique arguments)
$[+o]$ complements   ("internal arguments" of the predicate)
$[-r]$ core arguments   (subject and object only)
$[+r]$ non-core arguments   (all arguments except subject and object)

These features are associated with positions in a universally available lexical valency frame, from which predicates select a subset of argument positions:

(102)   ⟨ arg_1 arg_2 arg_3 ... arg_4 ...   ⟩
$[-o]/[-r]$ $[-r]$ $[+o]$ $[+o]$ $[-o]$ $[-o]$

The ordering and feature assignment in (102) is based on the standard LFG Functional Hierarchy, repeated in (103):

(103)   *The Functional Hierarchy:*
SUBJ > OBJ > OBJ$_\theta$ (> XCOMP, COMP) > OBL$_\theta$ (> XADJ, ADJ) .

The first position in (102), called mnemonically arg_1, corresponds to the canonical subject, and is associated with one of the two features which describe the SUBJ function (it is marked $[-o]$ in unergative predicates, emphasising its non-complement status, and $[-r]$ in unaccusative ones, emphasising its core status).[32] The second position, arg_2, corresponds to the canonical direct object, and is marked $[-r]$ (core). The next position, arg_3, corresponds to the restricted object, and is marked $[+o]$ (complement). Lastly, arg_4, corresponds to a canonical oblique argument, and is marked $[-o]$ (non-complement). Predicates can select any number of arguments from this frame, but, as indicated, they can only choose one arg_1 and arg_2, though they can select multiple arg_3s and arg_4s – this corresponds to the fact that a predicate can subcategorise for only a single SUBJ and OBJ, whereas multiple OBJ$_\theta$s and OBL$_\theta$s are permitted, being individuated by their subscripts (e.g. OBJ$_\text{THEME}$ vs. OBJ$_\text{BEN}$).[33]

What we have considered as mapping so far in this chapter corresponds to "argument-function mapping" in Kibort MT, i.e. the linking of argument positions and GFs. As in Classical LMT, arguments in Kibort MT are associated with

---

[32]Although the unergative/unaccusative distinction was originally applied only to intransitive predicates (Perlmutter 1978), subsequent work has extended it to predicates of all valencies: see Kibort (2004: 74–75) for discussion, and cf. the Dutch experiencer verbs discussed in Section 4.2.2, which exhibited the same syntactic split as intransitive unergatives/unaccusatives.

[33]While these functions are often indexed by thematic roles, this can be understood purely for distinctiveness, having no semantic content: instead of OBJ$_\text{THEME}$ and OBJ$_\text{BEN}$ we could use other mnemonic labels such as cases (e.g. OBJ$_\text{ACC}$ vs. OBJ$_\text{DAT}$, etc.) or preposition names (e.g. OBL$_\text{TO}$ vs. OBL$_\text{ON}$, etc.), or purely arbitrary labels such as OBJ$_1$ and OBJ$_2$. Thus, the retention of the GFs OBJ$_\theta$ and OBL$_\theta$ does not diminish the syntactically-motivated characterisation of GFs in Kibort MT.

a feature specification that makes them compatible with two different GFs, and mapping therefore consists in determining which of the two (if either) will realise the argument syntactically. Kibort MT diverges from Classical LMT, however, in only having a single Mapping Principle (Kibort 2014: 267; cf. Her 2013):

(104) *Mapping Principle (Kibort MT):*
       The ordered arguments are mapped in turn onto the highest (i.e. *least* marked) compatible grammatical function on the Markedness Hierarchy.

This inverts Mapping Principle (b) of Classical LMT, which maps arguments to the lowest, i.e. *most* marked, compatible GF, and in so doing removes the need for Mapping Principle (a), along with the Subject Condition, as we shall see. This is clearly a huge gain in parsimony, though it is not without cost, as we discuss below.

By way of illustration, consider again the simple transitive (and unergative) verb *kick*. This has the following Kibort MT a-structure:

(105) *kick* $\langle$ arg_1 arg_2 $\rangle$
                          $[-o]$ $[-r]$

By the Mapping Principle, we first map the highest argument, arg_1, onto the highest compatible GF: in this case, the highest $[-o]$ GF is subj, so this is what we choose. Next, arg_2 is mapped onto the highest $[-r]$ GF available: since subj is already taken, this is obj.[34] Note that despite the procedural talk here and in the Mapping Principle itself (arguments are mapped "in turn"), this process is intended to be understood declaratively. It can be seen as optimising the alignment between two hierarchies: are the highest arguments linked to the highest GFs? This can then be solved using various constraint-based tools such as those of Optimality Theory (Prince & Smolensky 1993, 2004; cf. also Asudeh 2001 for an application of OT to mapping in an LFG context).

Morphosyntactic argument alternations interfere with the default argument-function mapping. As in Classical LMT, this is achieved monotonically, by further specifying the mapping possibility of an argument. However, Kibort MT goes even further in this respect, eschewing the use of suppression altogether, and thus sidestepping the issues mentioned in fn. 22. For instance, Kibort (2001: 170) treats passivisation as a further specification of arg_1 as $[+r]$, illustrated in (106) for passive *kicked* (cf. (50) above):

---

[34]Function-Argument Biuniqueness still applies in Kibort MT, although it may not be necessary to stipulate it as a separate principle – see fn. 19.

(106)   *kicked*_{PASSIVE}   ⟨   arg$_1$ arg$_2$ ⟩
                                         [−o]   [−r]
                                         [+r]

The argument which by default would map to SUBJ is instead fully specified as an OBL$_\theta$, and, as a result, the arg$_2$, if there is one, becomes the SUBJ. Note that this gives the correct result for the English long passive, where the Agent is expressed as an oblique *by*-phrase, but in the short passive the Agent is not expressed grammatically at all. Kibort (e.g. 2004: 29) refers to such obliques as "optional", but it is not clear what determines this – it cannot be the case that OBL$_\theta$s are always optional, for instance, since there are certainly cases of obligatory obliques, as in *I gave the book *(to my friend)*.

In general, morphosyntactic operations are assumed to involve making arguments more marked, by adding additional +-valued specifications:

(107)   a. adding the [+r] specification to a [−o] argument
             (e.g. passivisation)

         b. adding the [+r] specification to a [+o] argument
             (e.g. secondary object preservation – Kibort 2007: 268)

         c. adding the [+o] specification to a [−r] argument
             (e.g. locative inversion – Kibort 2004: 364–367)

One thing to note about argument-function mapping in Kibort MT is that the Subject Condition of Classical LMT is absent. The motivation for this is that genuinely subjectless predicates are quite common in the world's languages (see Kibort 2006 and Lowe et al. 2021 for discussion). For instance, Polish intransitives can be passivised, resulting in a subjectless sentence (Kibort 2006: 304–307):

(108)   Było   codziennie sprzątane   (przez firmę).
         was.3SG.N every-day clean.PART.SG.N (by   company)
         'There was cleaning every day (by a company).'

This follows quite naturally in Kibort MT, where the verb will have the following a-structure, resulting in the first and only argument being mapped to OBL$_\theta$, rather than SUBJ:

(109)   *sprzątać*_{PASSIVE}   ⟨   arg$_1$ ⟩
                                           [−o]
                                           [+r]

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

The strong cross-linguistic preference for subjects is captured in the Mapping Principle: since arguments are mapped to the highest available GF on the Markedness Hierarchy, and since SUBJ is at the top of that hierarchy, SUBJ will always be the most preferred GF, meaning *something* will usually map to it. But by making this is a strong preference rather than a principle of the grammar, Kibort MT also allows for the possibility of subjectless predicates in marked circumstances – such as the passivisation of an intransitive.

One negative side effect of this choice, however, is that Kibort MT apparently makes the wrong predictions about the passive of double object verbs. As mentioned above in Section 4.1.3, when a double object verb is passivised, and so the primary object is promoted to SUBJ, it is apparently *not* the case that the secondary object is promoted to primary object – but this is exactly what Kibort MT predicts should happen, since the [+o]-valued arg_3 of a secondary object argument is compatible with OBJ, and OBJ is less marked than OBJ_θ (though see Kibort 2008).

The Kibort MT approach to argument-function mapping offers a different perspective from Classical LMT, and perhaps represents an advancement in certain areas, in particular with respect to theoretical parsimony. However, the real advantage of the theory is in the fact that argument-*participant* mapping can interact in interesting ways with argument-function mapping. Let us return now to the question of the *spray/load* alternation. The verb *spray* in this sense will have the following a-structure and argument-function mappings:

(110)   *spray*   ⟨ arg_1 arg_2 arg_4 ⟩
                    [−o]   [−r]   [−o]
                      | | |
                    SUBJ   OBJ OBL_θ

In fact, these GFs are the same ones which appear in both alternants – the only difference is which participants map to which GFs. Because Kibort MT posits a separate level of semantic participants, the mapping between those participants and the argument positions – and so, indirectly, the GFs – can be allowed to vary.

(111)   a. Adam sprayed the wall with the paint.

b.   *spray*   ⟨   Agent   Goal   Theme   ⟩

$$\text{b.} \quad spray \quad \langle \quad \underset{[-o]}{\underset{\mid}{arg_1}} \quad \underset{[-r]}{\underset{\mid}{arg_2}} \quad \underset{[-o]}{\underset{\mid}{arg_4}} \quad \rangle$$

	Agent	Goal	Theme
	$\mid$	$\mid$	$\mid$
b. *spray* ⟨	arg_1	arg_2	arg_4 ⟩
	$[-o]$	$[-r]$	$[-o]$
	$\mid$	$\mid$	$\mid$
	SUBJ	OBJ	OBL$_\theta$

(112)   a.   Adam sprayed the paint on the wall.

	Agent	Goal	Theme
b.	$\mid$	$\mid$	$\mid$
spray ⟨	arg_1	arg_2	arg_4 ⟩
	$[-o]$	$[-r]$	$[-o]$
	$\mid$	$\mid$	$\mid$
	SUBJ	OBJ	OBL$_\theta$
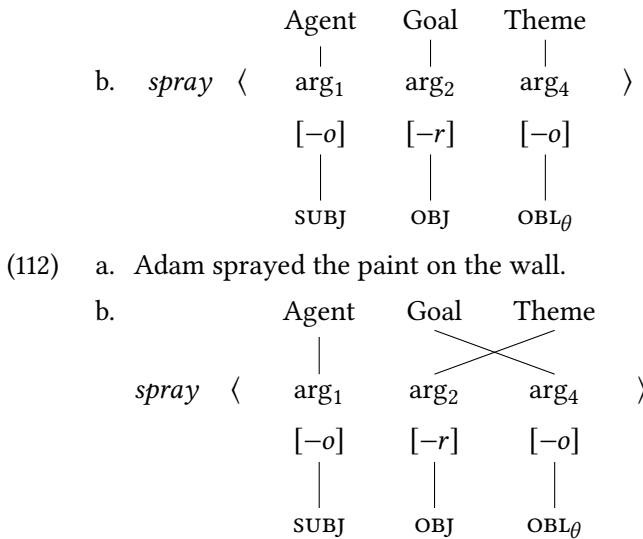
Although for a human reader it may be easier to track the re-aligned participants in diagrams like (111b) and (112b) if they are represented by thematic role labels, Kibort MT takes the criticisms of thematic roles mentioned in Section 2.1 to heart, and so they play no role in the theory. Furthermore, Kibort (2014) argues that neither Dowty-style proto-roles nor feature decomposition attempts are adequate either. In the absence of an adequate and complete representation of lexical knowledge, Kibort MT instead adopts a very minimal representation of semantic participants. In this system, semantic participants are labelled by numbers which identify which arg positions they can map to (Kibort 2014: 275ff.). For example, the a-structure of *spray* would be augmented as follows:

	1	24_{TH}	24_{GO}
(113) *spray* ⟨	arg_1	arg_2	arg_4 ⟩
	$[-o]$	$[-r]$	$[-o]$
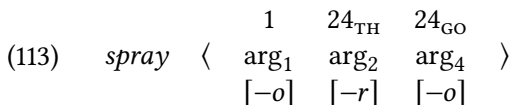
The first semantic participant is labelled 1 since it can only be linked to the arg_1 position, but the other two are labelled 24 since they can be linked to either the arg_2 or the arg_4 position. The subscripts on the semantic participants are purely for distinctness, to individuate the two participants with identical labels, and have no semantic content.

Argument-participant mapping has no principles beyond stating that participants with label *n* can be linked to argument arg$_n$; arguments whose labels contain multiple numbers, like the Theme and Goal in (113), are assumed to bear multiple labels, i.e. each of the Theme and Goal in (113) simultaneously has the

label 2 and the label 4. In cases where multiple mappings are possible, Kibort MT predicts that neither is more basic than the other, since there is no preference ranking encoded in the argument-participant mapping. This is certainly right for the *spray/load* alternation, since there does not seem any reason to assume that one alternant is derived from the other or that one is more basic than the other, especially given that this alternation is unmarked in English (i.e. there is no morphological or syntactic marker in either version).[35]

Kibort MT thus draws a clear formal distinction between morphosyntactic (meaning-preserving) and morphosemantic (meaning-altering) alternations: the former affect the argument-function mapping, using techniques very similar to those of Classical LMT; the latter affect the argument-participant mapping, something made possible by separating out these two levels of representation.

In sum, Kibort MT offers a mapping theory that on the one hand simplifies, and on the other hand elaborates on Classical LMT. It is simpler in that there is a universal valency frame, a single Mapping Principle, and no mention of thematic roles, but it is more complex in that it separates out the notion of argument from semantic participant. This does, however, offer the possibility of straightforwardly representing the effects of meaning-altering, morphosemantic alternations, something that was not always possible in Classical LMT.

6 Formal issues and recent developments

Aside from Kibort's focus on expanding the empirical coverage of LMT, another major thread in contemporary work on argument structure and mapping theory has been an increased interest in questions of formalisation. In this section, we address three areas in this vein: the formal status of a-structure, the nature of mapping, and the integration of mapping theory and compositional semantics.

6.1 The position and nature of a-structure

In Kibort MT, Classical LMT, and earlier work, the position of argument structure in the architecture of the grammar is left vague or unmentioned. Sometimes, it is (implicitly) assumed to be situated inside f-structure, as (part of) the value of PRED, but otherwise the question does not arise.

---

[35]It may be possible to argue that one of the variants is more basic on non-linguistic grounds, e.g. by reference to the relative prominence of cognitive concepts like Figure and Ground (Talmy 1978; see also Schätzle 2018 for an implementation of these concepts within LFG's mapping theory), but a strength of Kibort MT is that such a move is not *necessary*, even if it may sometimes be independently motivated.

Butt et al. (1997: 1) are the first to address this formal deficiency head on, and propose that argument structure forms its own level of representation, a-structure, situated in the LFG projection architecture between c-structure and f-structure:

(114)

$$V \xrightarrow{\alpha} \begin{bmatrix} \text{REL} & \text{eat} \\ \text{AGENT} & [\ ] \\ \text{THEME} & [\ ] \end{bmatrix} \xrightarrow{\lambda} \begin{bmatrix} \text{PRED} & \text{`eat'} \\ \text{SUBJ} & [\ ] \\ \text{OBJ} & [\ ] \end{bmatrix} \xrightarrow{\sigma} \begin{matrix} [\ ] \\ [\ ] \\ [\ ] \end{matrix}$$

$$\phi = \lambda \circ \alpha$$
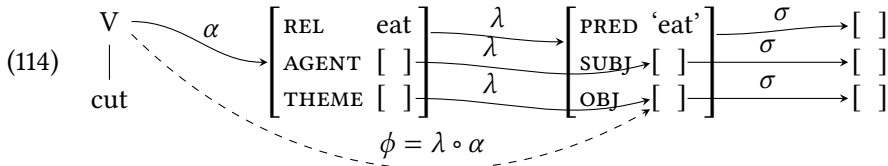
(Butt et al. 1997: 1, their ex. (1))

This positioning is motivated by the complex predicate facts discussed in Section 4.2.4. Since complex a-structures can correspond to simplex (monoclausal) f-structures, and since the projection functions, as functions, can be many-to-one but not one-to-many, a-structure must be mapped to f-structure, and not vice versa. On the other hand, since complex a-structures can be built from discontinuous pieces in the syntax, and are not necessarily generated in the lexicon, a-structure must be positioned after c-structure, so that information can be passed from the latter to the former.

One immediate effect of this positioning is to break up the traditional ϕ mapping from c- to f-structure: it is now the composition of two functions, the α function from c- to a-structure, and the λ function from a- to f-structure, i.e. $\lambda \circ \alpha$.[36] Some have seen this as undesirable: for example, Asudeh & Giorgolo (2012) propose a change to the architecture (to be discussed shortly), one of the effects of which is to restore ϕ to its atomic status, and they claim this as an advantage of their proposal (Asudeh & Giorgolo 2012: 71) – but if this is an advantage, we do not see how it can be anything other than an aesthetic one.

Unlike in most earlier approaches, for Butt et al. (1997), a-structures are not simply lists of arguments, but are instead AVMs. This allows for a richer internal structure: for example, complex predicates have nested a-structures (Butt et al. 1997: 12). Each a-structure contains a REL attribute that names the semantic relation it encodes, and attributes labelled with thematic role names corresponding to argument positions. Nothing further is said about the value of these attributes, and they are represented as empty AVMs in Butt et al. (1997). These must be shorthand for more complete structures, however, since otherwise, under a standard

---

[36]Butt et al. (1997: 1) identify ϕ with $\alpha \circ \lambda$ (rather than $\lambda \circ \alpha$), but this must be an error, since α has to be applied before λ, given their architecture.

set-theoretic interpretation of AVMs, all the "empty" AVMs would in fact be one and the same.[37]

Asudeh & Giorgolo (2012) criticise Butt et al.'s (1997) architecture and propose an alternative which has since proven influential. They do so on the basis of verbs which take optional objects, like *eat* in English:

(115) a. Donatello ate a pizza earlier.

b. Donatello ate earlier.

Although the Patient argument does not need to be expressed in the syntax, it must still be present in the a-structure, since it remains part of the core relation expressed by the verb (eating events involve something being eaten), and must also be represented at s-structure, since it is interpreted semantically: the truth of *Donatello ate* implies the truth of *Donatello ate something*. This poses a problem for the Butt et al. (1997) architecture, since there is no route through the projection architecture from the a-structure PATIENT to its corresponding s-structure without going via its f-structure representation, and it appears not to have one:

(116)
$$
\begin{bmatrix} \text{REL} & \text{eat} \\ \text{AGENT} & [\ ] \\ \text{PATIENT} & [\ ] \end{bmatrix}
\xrightarrow{\lambda}_{\lambda}
\begin{bmatrix} \text{PRED} & \text{`EAT'} \\ \text{SUBJ} & [\text{PRED 'DONATELLO'}] \\ \text{ADJ} & \{[\text{PRED 'EARLIER'}]\} \\ \text{TENSE} & \text{PAST} \end{bmatrix}
\xrightarrow{\sigma}_{\sigma}
\begin{array}{c} [\ ] \\ [\ ] \\ [\ ] \end{array}
$$
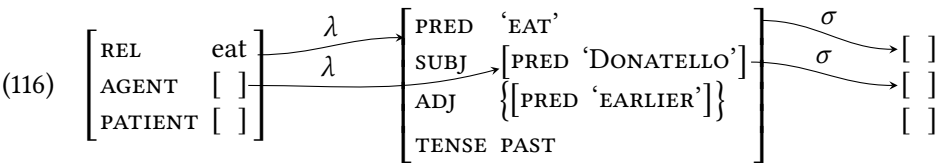
One might therefore be tempted to posit an unpronounced OBJ attribute at f-structure corresponding to the Patient, but there is empirical evidence against this (Asudeh & Giorgolo 2012: 71). For example, this putative null pronoun cannot antecede another, subsequent pronoun:

(117) a. Donatello ate a pizza, but it turned out to be Raphael's.

b. * Donatello ate, but it turned out to be Raphael's.

Given this, we are forced to propose a new function which projects directly from a-structure to s-structure (i.e. it is not simply the composition of σ and λ); Asudeh & Giorgolo (2012: 70) call this the θ projection. (118) shows this new situation:.

---

[37]For discussion of a similar problem, this time with regards to s-structure, see Findlay (2021: 348–353).

(118)

$$
\begin{bmatrix} \text{REL} & \text{eat} \\ \text{AGENT} & [\ ] \\ \text{PATIENT} & [\ ] \end{bmatrix}
\xrightarrow{\substack{\theta \\ \lambda \\ \lambda \\ \theta \\ \theta}}
\begin{bmatrix} \text{PRED} & \text{`EAT'} \\ \text{SUBJ} & [\text{PRED `DONATELLO'}] \\ \text{ADJ} & \{[\text{PRED `EARLIER'}]\} \\ \text{TENSE} & \text{PAST} \end{bmatrix}
\xrightarrow{\substack{\sigma \\ \sigma}}
\begin{matrix} [\ ] \\ [\ ] \\ [\ ] \end{matrix}
$$
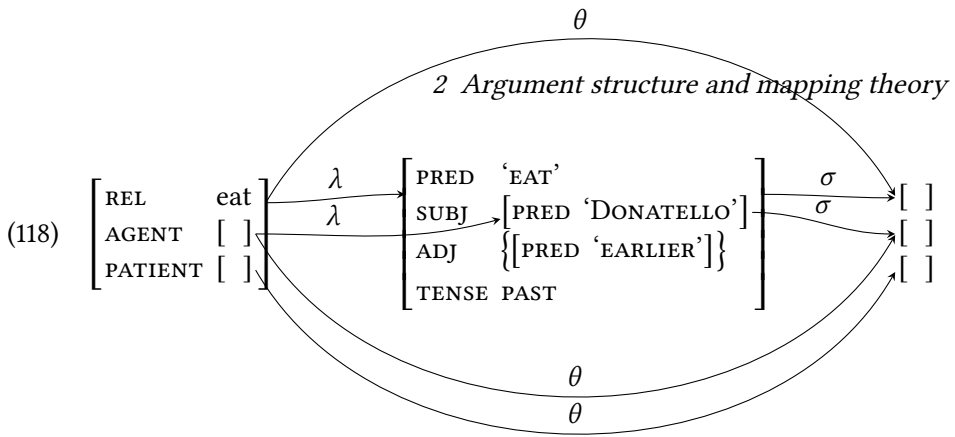
This move adds formal complexity to the grammar (a whole new projection function) and also adds indeterminacy: when an element of a-structure *is* expressed at f-structure, there are now two ways of reaching its s-structure – one via $\sigma \circ \lambda$ and one via θ directly. Even if this solves the problem of unexpressed arguments, it is a formally unhappy scenario to be forced into.

Asudeh & Giorgolo's (2012) solution is to do away with a-structure as a separate level of representation, and to replace it with a new, connected version of s-structure – that is, rather than the s-structures for the arguments being separate from the s-structure for the clause (and from each other), they are instead embedded inside it. This makes this new conception of s-structure very similar to Butt et al.'s (1997) a-structures. An example is shown in (119):

(119)

$$
\begin{bmatrix} \text{PRED} & \text{`EAT'} \\ \text{SUBJ} & [\text{PRED `DONATELLO'}] \\ \text{ADJ} & \{[\text{PRED `EARLIER'}]\} \\ \text{TENSE} & \text{PAST} \end{bmatrix}
\xrightarrow{\substack{\sigma \\ \sigma}}
\begin{bmatrix} \text{REL} & \text{EAT} \\ \text{ARG}_1 & [\text{REL DONATELLO}] \\ \text{ARG}_2 & [\text{REL VAR}] \end{bmatrix}
$$
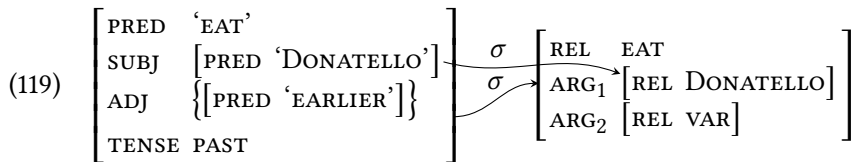
Ultimately, it is a fairly arbitrary choice whether we call this new connected structure s-structure or a-structure. Asudeh & Giorgolo (2012) call it s-structure since they continue to use it as part of the linear logic component of Glue Semantics meaning constructors, but it has a lot in common with Butt et al.'s (1997) a-structure as well, being internally structured/connected and expressing the predicate-argument structure of the clause. What is more, later developments have sought to imbue this new structure with additional information about tense, aspect, and event structure (see e.g. Lowe 2014, Lovestrand 2018, 2020, Findlay 2021), thereby incorporating some information which is also present in Butt's (1995) "elaborated" a-structures (on which see below). For consistency with other work, however, we will continue to call these s-structures here.

The exact content of these s-structures is subject to ongoing research, but they are assumed to at least include a REL attribute identifying the semantic relation

expressed (cf. Asudeh et al. 2013: 24), and potentially several numbered ARG attributes, e.g. ARG$_1$, ARG$_2$, for each of that relation's arguments. Asudeh & Giorgolo (2012) use REL only for predicates, and leave argument s-structures as "empty" AVMs, just like Butt et al. (1997). Lovestrand (2018: ch. 8.3) and Findlay (2020: 135f.), however, generalise the presence of REL to argument as well as predicate s-structures, and Findlay (2020: 144) proposes to use "var" as the REL value for unexpressed/suppressed arguments.

The numbered ARG attributes are used instead of Butt et al.'s (1997) thematic role labels in part because Asudeh & Giorgolo (2012) make use of a neo-Davidsonian meaning language (Parsons 1990) such that thematic role information is expressed directly in the semantics – i.e. instead of (120a), the meaning of *eat* is expressed by (120b) – and so it would be redundant to also encode this information in s-structure.

(120)   a. $\lambda x \lambda y \lambda e.\textbf{eat}(e, x, y)$
         b. $\lambda x \lambda y \lambda e.\textbf{eat}(e) \wedge \textbf{agent}(e, x) \wedge \textbf{theme}(e, y)$

This has the additional benefit of relegating thematic roles to the meaning language rather than making them part of the meta-language of the grammar itself. There they can be treated as abbreviations for whatever sets of semantic entailments we take them to encode (*à la* Dowty 1991), with whatever level of granularity is required, leaving the grammar itself free of the nebulous notion of thematic role.

The significance, or lack thereof, of the ARG labels has been the subject of disagreement, however. They were originally intended as arbitrary labels merely to achieve distinctness at s-structure, but Findlay (2016) imbues them with meaning, identifying them with the numbered arg positions of Kibort MT (see Section 5), as part of an implementation of that theory within the new architecture. This view has been adopted by others (e.g. Asudeh et al. 2014, Lowe 2016, Lovestrand 2018, 2020), but Findlay (2020) argues for a return to the *status quo ante*, where these labels have no significance in and of themselves, and shows that the same implementation of Kibort MT can be achieved while avoiding reifying the s-structure attribute names.

The title of Findlay (2016) is "Mapping theory without argument structure", but this is in many respects a mischaracterisation of the research programme inspired by Asudeh & Giorgolo's (2012) architectural proposal. Rather than doing away with argument structure, this work has served more as a rationalisation of the LFG architecture: instead of having two levels, a-structure and s-structure,

the latter of which is rather informationally impoverished, we have a single level of representation which shares properties of both.[38]

As mentioned above, some researchers have imbued this new structure with additional information about lexical semantics and event structure (e.g. Lowe 2014, Lovestrand 2018). But suggestions to add this kind of information to a-structure are not new. Butt (1995) develops what she calls an ELABORATED A-STRUCTURE (Butt 1995: 133), which includes much more structure and much more semantic information than Classical LMT's minimalist a-structures. This elaborated a-structure is based on Jackendoff's (1990) LEXICAL CONCEPTUAL STRUCTURES (LCSs), but only includes the concepts relevant to linking and semantic case marking (Butt 1995: 143). An example of the elaborated a-structure for the Urdu main verb *de* 'give' is shown in (121):

$$
(121) \quad
\begin{bmatrix}
\text{de 'give'} \\
\begin{bmatrix}
\text{CS}([\alpha], \text{GO}_{Poss}([\ ], \text{TO}[\ ])) \\
\text{AFF}([\ ]^{\alpha},) \\
\text{ASP}(\_\ \_\ \_)
\end{bmatrix}_{E}
\end{bmatrix}
$$

The inner box is the actual a-structure, and contains three levels. The first two are borrowed from Jackendoff's LCSs: the THEMATIC TIER and the ACTION TIER. The former, the Thematic Tier, describes the lexical meaning of the verb in decompositional terms – here that one entity *causes* (CS) *possession* of another to *go* (GO$_{Poss}$) *to* a third entity (TO). The latter, the Action Tier, describes the relationship between Actor, Patient, and Beneficiary roles – in other words those roles

---

[38]The observant reader may be entertaining an architectural concern at this point: earlier, we motivated the Butt et al. (1997) architecture by drawing on the facts of complex predicates: a complex a-structure can correspond to a simplex (monoclausal) f-structure, and so we need the former to precede the latter in the projection architecture in order to retain the functional nature of the projection relations. However, in the new architecture, the connected s-structure which represents predicate-argument structure comes after f-structure, so we appear to be in trouble. Two solutions to this puzzle have been proposed. Lowe (2016) gives the first analysis of complex predicates in this new framework, and argues that they should be given a flat s-structure (in contrast to the articulated a-structures usually assumed), representing their complexity in the meaning language instead. This avoids any problems arising from having a flat f-structure, since it is no longer required to subsequently project a more articulated s-structure. Alternatively, Lovestrand (2020) proposes to give complex predicates articulated f-structures after all, which means a complex s-structure is also possible without losing the functional nature of σ. There are empirical shortcomings with both of these approaches, but they fare no worse than existing, alternative approaches, and serve to illustrate how the apparent monoclausality of complex predicates does not force us to assume an articulated a-structure which precedes f-structure in the projection architecture.

which usually receive structural case. As Butt (1995: 137) points out, it can also be thought of as encoding an analogue of Dowty's (1991) proto-roles. Here the argument labelled α, i.e. the "giver" (the one causing the transfer of possession) is indicated to be *affecting* (AFF) something else. The second slot of the function AFF is left empty, indicating that there is no true Patient or Beneficiary here (Butt treats the recipient as a simple Goal instead of a Beneficiary). There are also subtypes of the AFF function which provide information about volitionality or conscious choice.

The final tier is the Aspect Tier. This is not borrowed from Jackendovian LCSs, but is an innovation by Butt. It represents aspectual information: specifically, whether a verb is positively or negatively specified for inception, duration, and/or completion (Butt 1995: 142). The function ASP contains three slots, one for each of these properties, and each can be specified positively, with a '1', negatively, with a '0', or left unspecified, indicated by a '_'. In (121), all three slots are empty, showing that this verb is unspecified for this aspectual information.

Clearly, this conception of argument structure is far more complex than the ordered lists used in Classical LMT, and more informationally rich than either of the structures discussed already in this section.[39] Butt argues that this complexity is motivated by its capacity to offer an elegant account of complex predicates. For one thing, the elaborated a-structures expose more lexical semantic content to the grammar, enabling appropriately fine-grained constraints to be placed on complex predicate formation (see e.g. Butt 1995: 147–155 for examples). For another, they add articulation and structure, and, as we saw in Section 4.2.4, the proper treatment of complex predicates necessitates assuming a more articulated a-structure than is standard in Classical LMT – at least one capable of recursive embedding.

On Butt's (1995) approach, the light verbs which are used in complex predicates have a-structures which themselves have argument slots for *other a-structures*, labelled as transparent events (E_T), since the light verbs that host them can "see into" their internal structure. This visibility allows different kinds of argument fusion to take place, whereby participants of the embedded event are identified

---

[39]Indeed, one reviewer suggests that the level of representation proposed by Butt (1995) is not argument structure at all, but rather some kind of "event structure" or "semantic structure". To the extent that the additional information is necessary to handle argument structure phenomena like complex predicate formation, and given that these structures also do everything else we would want from an argument structure (see e.g. Butt 1995: ch. 6 on mapping), it is hard to know what to make of this complaint. Perhaps a more minimal a-structure would in fact be sufficient, but if so that is a matter to be demonstrated empirically, rather than settled by definitional fiat.

with participants of the event described by the light verb (as discussed in Section 4.2.4). We omit the full details here – see Butt (1995: ch. 5) for more information. By way of illustration, the a-structure for the Urdu permissive light verb *de-* 'let' is given in (122) (Butt 1995: 156):

$$(122) \quad \begin{bmatrix} \text{de- 'let'} \\ \begin{bmatrix} \text{CS}([\alpha], \text{GO}_{Poss}(\{\ \}_{E_T}, \text{TO}[\ ])) \\ \text{AFF}([\ ]^\alpha,) \\ \text{ASP}(\_\ \_\ \_) \end{bmatrix}_E \end{bmatrix}$$

This is very similar to the a-structure in (121), the only difference being that the first argument of GO_{Poss} has been replaced by a transparent event (indicated by the curly braces and subscript E_T). The "letting" event expressed by this light verb is viewed metaphorically as a transfer event, where the thing transferred is the permitted event. This gives some explanation to the fact that both verbs share the same form in Urdu, for example, and shows how the embedded verb contributes to the overall interpretation of the complex predicate. It also allows for the recursive construction of complex predicates which are embedded under more than one light verb.[40]

A more contemporary approach to expanding the coverage of a-structure, but without assuming the Asudeh & Giorgolo (2012) architecture, is that of Schätzle (2018: ch. 6). She assumes a richly multidimensional version of Kibort MT's a-structure, where each argument can be annotated with a variety of non-standard semantic information, such as whether it is a FIGURE or GROUND (Talmy 1978), and which kind of event participant it is in the typology of Ramchand's (2008) FIRST-PHASE SYNTAX. This, Schätzle (2018: 202) claims, enables a more "semantically realistic" account of mapping and of argument alternations, a goal shared by other recent work – see Section 6.3.

6.2 Mapping as co-description

The relationship between different levels of structure, such as a-structure and f-structure, has been approached in two different ways in LFG: CO-DESCRIPTION and DESCRIPTION BY ANALYSIS (Kaplan 1995, Dalrymple et al. 2019: 267–270). In

---

[40]Other work on complex predicates and LMT, including Butt's own later work, has tended to eschew these more complex a-structures in favour of the simpler, ordered list representations of Classical MT (e.g. Alsina 1996, 1997, Butt 2014). But this leads to enormous difficulty in appropriately formalising the process of PREDICATE FUSION: see Lowe (2016: sec. 2) for critical discussion.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

co-description, multiple levels of structure are described simultaneously – for example, LFG's annotated phrase-structure rules simultaneously describe both c-structure and f-structure. This is the most commonly used approach in LFG. The alternative, description by analysis, involves determining the description of one structure by inspecting and analysing another. This was used in early LFG proposals for semantic analysis (e.g. Halvorsen 1983). Findlay (2021: 344–345) discusses various shortcomings of the description by analysis approach: notably, it ignores the possibility of mismatches between levels, and fails to meet the desideratum of constraint-based grammars laid down by Pollard & Sag (1994: 13) that they be "process neutral": description by analysis inevitably introduces directionality into parsing, which co-description does not. Co-description therefore "most directly captures the spirit of the constraint-based approach to linguistic analysis" (Findlay 2021: 344), which may explain why it has come to dominate in LFG analyses – indeed, while description by analysis was prominent in early accounts of semantics in LFG, those approaches have since been replaced by Glue Semantics (Dalrymple, Lamping & Saraswat 1993, Dalrymple 1999, Asudeh 2022), which employs co-description.

Classical LMT, though, is very much in the spirit of description by analysis: GF assignments at f-structure are determined by inspecting a-structure, and by analysing it using the Mapping Principle(s). This state of affairs meant that LFG work on argument structure and mapping was out of sync with the theoretical mainstream, where co-description was the norm. Once again, Butt et al. (1997: 6) were the first to tackle this formal issue, treating mapping as co-description of both a- and f-structure.

For example, to say that a predicate's Agent argument is expressed as its SUBJ GF, we could include the following piece of functional description in its lexical entry (where $*$ refers to the c-structure node bearing the annotation, and $\hat{*}$ to its mother node):

(123) $(\hat{*}_\alpha \text{ AGENT})_\lambda = (\hat{*}_{\alpha\lambda} \text{ SUBJ})$

The expression $\hat{*}_\alpha$ refers to the lexical item's a-structure, via the α projection from c- to a-structure, while the expression $\hat{*}_{\alpha\lambda}$ refers to the lexical item's f-structure (the equivalent of the more familiar ↑). This constraint therefore picks out the f-structure corresponding to the a-structure AGENT, and identifies it with the verb's f-structure SUBJ.

But, of course, we generally don't want to associate an argument with only a single GF. Instead, Classical LMT associates it with a feature which describes a *pair* of GFs. Butt et al. (1997: 6) make this disjunctive meaning of the features

explicit: instead of associating an argument with a feature, a disjunction of mapping equations like (123) is given, as in (124) or (125):

(124) AGENT *links to* $[-o]$:
$(\hat{\ast}_\alpha \text{ AGENT})_\lambda = (\hat{\ast}_{\alpha\lambda} \text{ SUBJ}) \vee$
$(\hat{\ast}_\alpha \text{ AGENT})_\lambda = (\hat{\ast}_{\alpha\lambda} \text{ OBL}_\text{AGENT})$

(125) THEME *links to* $[-r] \vee [+o]$:
$(\hat{\ast}_\alpha \text{ THEME})_\lambda = (\hat{\ast}_{\alpha\lambda} \text{ SUBJ}) \vee$
$(\hat{\ast}_\alpha \text{ THEME})_\lambda = (\hat{\ast}_{\alpha\lambda} \text{ OBJ}) \vee$
$(\hat{\ast}_\alpha \text{ THEME})_\lambda = (\hat{\ast}_{\alpha\lambda} \text{ OBJ}_\text{THEME})$

Butt et al. (1997: 6) suggest that these intrinsic specifications can be universal, like (124) for AGENTs and (125) for THEMEs, or they can be parameterised on a language-by-language basis, as is the case for other roles like LOCATION, GOAL, or INSTRUMENT.

Of course, these specifications alone do not determine the final mapping. In fact, Butt et al. (1997: 6) propose an important theoretical break from Classical LMT in this respect:

> Our approach departs most radically from the LMT literature in that we do not assume that a-structure roles are deterministically and uniquely linked to grammatical functions via a set of default principles. Instead, we propose a set of preference constraints which impose an ordering on the available linking possibilities; the most preferred possibility or possibilities are chosen.

In essence, their approach rejects the mechanistic, rule-driven approach of Classical LMT, and instead proposes that there is a hierarchy of GFs, and that those mappings which realise more highly ranked GFs are preferred. The hierarchy they propose is as follows:[41]

---

[41]Butt et al. (1997: 7) claim that the hierarchy in (126) can be recast as a preference for negative-valued features in the classic $[\pm o/r]$ schema:

(i) a. $[-r] > [+r]$
    b. $[-o] > [+o]$

But the expressions in (i), which is their (15), do not match the authors' prose description, which only applies (i-b) within the $[-r]$ GFs. If we simply take (i) as expressing two independent preference rankings, we get the Markedness Hierarchy of Classical LMT (see Section 4.1.1):

(ii) SUBJ > OBJ, OBL$_\theta$ > OBJ$_\theta$

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

(126)   SUBJ > OBJ > OBL$_\theta$, OBJ$_\theta$

That is, SUBJ outranks OBJ, which in turn outranks OBL$_\theta$ and OBJ$_\theta$, which have the same rank as each other. This means, for each argument, that it is preferable for it to be realised as a SUBJ, or, failing that, as an OBJ, or, lastly, as either an OBL$_\theta$ or an OBJ$_\theta$. The argument will therefore be linked to the highest GF on this hierarchy with which it is compatible, given the disjunctive specifications provided in its intrinsic classification.[42] This gives us a much more dynamic system than in Classical LMT: there are no explicit Mapping Principles, and arguments simply compete for the highest available GFs. In a nod to Mapping Principle (a-i) of Classical LMT (see Section 4.1.2), Butt et al. (1997: 6) do include a preference for the SUBJ to be linked to the highest available argument on the thematic hierarchy, but crucially this is just a preference, and so is not inviolable.

The final mapping chosen is the one deemed "optimal" in terms of realising the highest number of the most highly ranked GFs, and in terms of satisfying any other preference constraints, such as the subject preference just mentioned (as well as not violating Function-Argument Biuniqueness or the Subject Condition). Butt et al. (1997: 7) use a numerical system to express the relative weightings of different GFs and of other constraints, but this is not a crucial component of the theory, and any appropriate means of ranking different solutions in terms of a set of preferences could be used – for example, the authors speculate (p. 7) that the proposal could be reformulated in terms of Optimality Theory (Prince & Smolensky 1993, 2004, *et seq.*).

By way of illustration, consider a simple transitive like *kick* again. For every argument, the most preferred GF is SUBJ. But is each compatible with SUBJ? According to the disjunctions in (124) and (125), assuming that the intrinsic classification for Theme also applies to Patients, SUBJ is a possible realisation of both arguments. But we cannot map both to SUBJ, or we fall foul of Function-Argument Biuniqueness, so we must decide which one to map to SUBJ, and which to map to the next most highly ranked compatible GF. Since, following the thematic hierarchy, the Agent argument of *kick* outranks its Patient argument, the subject

---

Alternatively, if we see (i-a) as taking precedence over (i-b), then we obtain another ranking, this time a total ordering:

(iii)   SUBJ > OBJ > OBL$_\theta$ > OBJ$_\theta$

It is of course an empirical matter which of these rankings (if any) is correct.

[42]Just like Kibort MT's Mapping Principle (see Section 5), this reverses the Classical LMT mapping principle where GFs *lower* down the hierarchy are preferred. This means that Butt et al.'s (1997) proposal shares the weakness of Kibort MT that it makes the wrong prediction about the passives of ditransitives – see Section 4.1.3.

preference will be satisfied if we map the Agent to SUBJ but not if we map the Patient to SUBJ, so the former mapping is preferred; the next highest GF compatible with the Patient intrinsic specification is OBJ, and so we end up with the correct outcome whereby the Agent is linked to SUBJ and the Patient to OBJ.

The theoretically most interesting consequence of the Butt et al. (1997) approach to mapping is that certain constructions may have more than one optimal linking. Butt et al. (1997: 8ff.) argue that this in fact characterises alternations which are motivated by semantic/pragmatic constraints (such as the dative shift) and not by morphosyntactic ones (such as the passive).[43] This offers a more natural account of the dative shift alternation than the Classical LMT analysis, which requires two different initial assignments of features to the arguments. In the Butt et al. (1997) framework, both realisations of the dative shift alternation in English are made available automatically, since they have equivalent preference rankings:

(127)   [Garak] gave [the datarod] [to Sisko].
        SUBJ   OBJ OBL_{GOAL}

(128)   [Garak] gave [Sisko] [the datarod].
        SUBJ   OBJ OBJ_{THEME}

Both involve a SUBJ (linked to the highest argument) and an OBJ, and since OBJ_θ and OBL_θ are equally ranked, the different realisations of the third argument make no odds when it comes to the relative weightings of the two mappings. Therefore both mappings are made available by the grammar, and the choice between them must be determined by other factors, such as lexical preference (the shifted variant is impossible with verbs of Latinate origin, for example) or semantic/pragmatic considerations (see Bresnan 2007 and Bresnan et al. 2007 for usage-based/probabilistic accounts of the alternation, and Goldberg 1995: ch. 6 on the special meanings associated with the double object construction in English).

Work which assumes the Asudeh & Giorgolo (2012) architecture also makes use of co-description to express mapping possibilities, although here the directionality is changed: we are mapping from f-structure to s-structure, rather than from a-structure to f-structure. The equivalent of (123), assuming ARG_1 corresponds to the Agent (see Section 6.3), is (129):

---

[43]However, their distinction does not seem to perfectly match that between meaning-preserving (morphosyntactic) and meaning-altering (morphosemantic) alternations, since they consider the locative inversion to be grouped with the dative shift (as being explained by the presence of more than one optimal linking) and distinct from the passive, when the locative inversion is no more meaning altering than the passive (neither alternation affects truth-conditional semantics, but only alters the information structural prominence of its arguments).

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

(129) $(\uparrow \text{SUBJ})_\sigma = (\uparrow_\sigma \text{ARG}_1)$

As in Butt et al. (1997), feature decomposition is replaced by explicit disjunctions over GFs. Findlay (2016: 299) uses abbreviations to describe the (supposedly) natural classes captured by the traditional features:

(130)   a.   MINUSO   ≡ $\{\text{SUBJ}|\text{OBL}_\theta\}$
        b.   PLUSO   ≡ $\{\text{OBJ}|\text{OBJ}_\theta\}$
        c.   MINUSR   ≡ $\{\text{SUBJ}|\text{OBJ}\}$
        d.   PLUSR   ≡ $\{\text{OBL}_\theta|\text{OBJ}_\theta\}$

This gives us (131) as the equivalent of (124):

(131) $(\uparrow \text{MINUSO})_\sigma = (\uparrow_\sigma \text{ARG}_1)$

In fact, since arguments may not be realised by any GF – for example, the Agent argument of a short passive – we also need a description which says that the argument in question does not correspond to any GF at f-structure. We achieve this by stating that the inverse of the σ mapping from f- to s-structure is empty when applied to that argument, as in (132):

(132) $(\uparrow_\sigma \text{ARG}_1)_{\sigma^{-1}} = \varnothing$

This says that the s-structure ARG_1 has no f-structure correspondent, i.e. that this argument is not realised syntactically.

Findlay (2016: 319,321) proposes to use templates to abbreviate these mapping equations and make them more readable:[44]

(133)   MAP(D,A)   ≡ $(\uparrow \text{D})_\sigma = (\uparrow_\sigma \text{A})$

(134)   NOMAP(A)   ≡ $(\uparrow_\sigma \text{A})_{\sigma^{-1}} = \varnothing$

The first of these, (133), says that the GF or disjunction of GFs D is mapped to the s-structure argument A, while (134) says that the s-structure argument A has no GF correspondent at f-structure.[45] These templates can then be combined, so that e.g. the correct expression to capture the mapping possibilities of an Agent assigned to ARG_1 is the following:

---

[44]On templates, see Dalrymple et al. (2004), Crouch et al. (2011), Asudeh et al. (2013) and Belyaev (forthcoming[a]: 77ff. [this volume]) in this volume.

> Check page numbering

[45]One problem with the NOMAP template is that in the event an argument is not expressed syntactically, nothing will ensure its presence at s-structure. Findlay (2020: 135–136) argues therefore that existential constraints must accompany the introduction of each argument.

(135) $\{@\textsc{Map}(\textsc{minuso}, \textsc{arg}_1) \,|\, @\textsc{NoMap}(\textsc{arg}_1)\}$

That is, either this argument is mapped to one of the two MINUSO GFs (SUBJ or OBL$_\theta$), or it is not expressed syntactically at all.

Using disjunctions over GFs like MINUSO or PLUSR instead of assuming features like $[-o]$ and $[+r]$ sidesteps any formal issues arising from seeing GFs as decomposable into features (as discussed in Section 4.1.1), and simply represents the most significant empirical claim of the feature-based approach – that GFs can be grouped into natural classes (whether the $[\pm o/r]$ classification is the correct way of grouping them is orthogonal). It has been objected that this use of disjunctions makes the approach somehow more arbitrary or less well motivated than earlier incarnations of LMT, since we could have just as easily written a different set of disjunctions in (130). Such an objection is misplaced for two important reasons. Firstly, it purports to contrast the arbitrariness of the disjunctive approach with the theoretical motivation of the feature-decomposition approach. But this is only true to the extent that the features used in the latter have independent motivations. While a case could be made for $[\pm r]$ on these grounds (one could imagine an independent criterion for determining semantic restrictedness), as we mentioned in Section 4.1.1, this seems not to be the case for $[\pm o]$, which has no content other than identifying the two object functions OBJ and OBJ$_\theta$, and whose definition is therefore circular. Given this situation, we take the use of the explicitly "arbitrary" mechanism of disjunction to in fact be an advantage over the classical approach, since it wears its arbitrariness on its sleeve rather than concealing it behind a veneer of theoretical motivation.

Secondly, and much more significantly, such an objection misses the crucial distinction between formalism and theory. The formalism *itself* need not be expected to say anything about what natural groupings of GFs occur in the world's languages. Rather, the formalism gives us tools for making explicit claims about such things – and it is those claims which constitute the theory. As Pollard (1997: 9) puts is, "it is the theory that imposes the constraints, not the language in which the theory is expressed". So, although we could've written different disjunctions in (130), it is precisely in writing one set of expressions rather than another that we make a theoretical claim. This claim may turn out to be true or false, but if it is false, we would prefer to be able to use the same familiar tools to express a different, revised hypothesis, rather than have to throw away our tools entirely because they have been over-engineered to fit one particular view of reality. Once again, therefore, we see this property as being an advantage of the disjunctive approach. As an example, consider the objection by Alsina (1996: 29, fn. 9), noted in Section 4.1.1, that the traditional $[\pm o/r]$ features cannot be used to describe the

natural class of terms, or direct GFs, i.e. SUBJ, OBJ, and OBJ$_\theta$. He instead proposes a different classification using the features [±subj/obl], where [−obl] describes the terms (Alsina 1996: 27–30). In the traditional view, this approach and the Classical LMT approach are simply incommensurable: they represent two different formalisms which contain different primitive elements. But in the view we are considering, both can be expresed in the *same* terms – compare (130) and (136) – thereby highlighting their status as competing theoretical claims rather than totally distinct formal approaches.

(136)   a.   MINUSSUBJ   ≡   {OBJ|OBJ$_\theta$|OBL$_\theta$}
        b.   PLUSSUBJ   ≡   SUBJ
        c.   MINUSOBL   ≡   {SUBJ|OBJ|OBJ$_\theta$}
        d.   PLUSOBL   ≡   OBL$_\theta$

It is an empirical matter which of these analyses is correct, and we should not generally expect the formalism to adjudicate on empirical matters. Rather, the theory which we develop in using that formalism is what we expect to align with the facts.

6.3 Connection to semantics

While the most influential research in Classical LMT was being conducted, there was no canonical theory of the syntax-semantics interface in LFG to appeal to. With the acceptance of GLUE SEMANTICS (Glue) into the LFG mainstream around the turn of the millennium, this changed.[46] One of the most important goals of recent work on mapping theory has therefore been to integrate the theory into a Glue-based analysis of the syntax-semantics interface. In particular, this strand of research assumes that Glue's concept of RESOURCE SENSITIVITY (Asudeh 2012: ch. 5) subsumes the traditional LFG principles of Completeness and Coherence, so that PRED features at f-structure no longer contain an argument list. That is, instead of (137a), we have (137b):[47]

---

[46]Although Glue first appeared in the early '90s (Dalrymple, Lamping & Saraswat 1993), it was still not well established in the LFG community by the time much of the the work discussed in the earlier sections of this chapter was carried out. The first major collection of Glue work connected to LFG was Dalrymple (1999), and the theory later appeared in Dalrymple's (2001) handbook-style presentation of LFG, as well as the latest reference guide to LFG, Dalrymple et al. (2019: ch, 8.5). We cannot include an introduction to Glue Semantics in this chapter for reasons of space, but see the references just cited, along with Asudeh (2022) and Asudeh (forthcoming) [this volume] for further information.

[47]The idea of using linear logic's resource sensitivity to account for Completeness and Coherence goes back to the very first Glue paper (Dalrymple, Lamping & Saraswat 1993), and was noted

mapping in Kibort MT (see Section 5). The crucial advantage of incorporating a theory of the syntax-semantics interface is that we can also express the equivalent of Kibort MT's argument-participant mapping, via the meaning constructor in the final line. Here the variable x is identified as the Agent of the kicking event, and connected via the linear logic term to ARG$_1$ at s-structure; similarly, y is identified as the Patient, and connected to ARG$_2$. That is, the link between GFs and semantic participants, a key part of any mapping theory, is mediated by the intervening level of s-structure, here playing the same role as Kibort MT's lexical valency frame. And just like in Kibort MT, this setup allows for the re-alignment of participants to argument positions – see Findlay (2016: 328–332) for an example of this with the English benefactive.

By bringing together information about mapping and about semantics, which are just the same kind of object in this approach, *viz.* pieces of functional description, it becomes far easier to express semantic constraints on, and semantic consequences of, argument alternations and other argument structure operations (cf. also the discussion of Butt's 1995 enhanced a-structures above). Asudeh (2021: 32–39) shows the potential of this approach in his analysis of the English "non-agentive dynamic intransitive", and contrasts it with what he calls the "low resolution" of Classical LMT, which only has access to very spartan semantic information (usually just the thematic roles of arguments).

One promising area of research made possible by this "joined up" approach to mapping is the idea of incrementally bundling up semantic and mapping information into more and more complex valency templates (as employed in e.g. Asudeh & Giorgolo 2012, Asudeh et al. 2014, Findlay 2020), which, coupled with the notion of an inclusion hierarchy between templates (see especially Asudeh et al. 2013: 17–20), could lead to a mapping theory based purely on a richly structured and hierarchical lexicon, along the lines of Davis & Koenig (2000). This potential has yet to be fully explored, though Przepiórkowski (2017) has pointed the way.

7 Conclusion

New approaches to argument structure and mapping theory phenomena were at the heart of what gave rise to LFG as a separate approach to linguistic theory in the first place: Bresnan's (1980, 1982) observations about the lexical character of argument alternations and the benefits afforded by separating out lexical predicate-argument structures from surface syntactic structures were what laid the foundations for LFG's lexicalist, modular view of the grammar. The advent

of Lexical Mapping Theory (LMT) helped to constrain the theory of argument alternations, and also offered new explanatory tools which proved successful in characterising a number of linguistic phenomena across a fairly typologically diverse range of languages. Recent developments in both theory and formalism show that the field is ripe for a renaissance, and that while great strides have been made, many important questions still remain unanswered. This chapter has attempted to give a broad and expository overview of the status quo, along with a little of how we got here, with the hope that by drawing together different theoretical perspectives we can both encourage dialogue among experienced researchers, and bring new scholars up to speed, so that both can be in the best position to contribute to a field which remains full of untapped potential.

Acknowledgements

References

Ackerman, Farrell. 1991. Locative alternation vs. locative inversion. In Aaron L. Halpern (ed.), *Proceedings of the 9th West Coast Conference on Formal Linguistics*, 1–13. Stanford: CSLI Publications.

Ackerman, Farrell. 1992. Complex predicates and morpholexical relatedness: Locative alternation in Hungarian. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical matters* (CSLI Lecture Notes), 55–83. Stanford: CSLI Publications.

Ackerman, Farrell & John Moore. 2001. *Proto-properties and grammatical encoding*. Stanford: CSLI Publications.

Ackerman, Farrell & John Moore. 2013. Proto-properties in a comprehensive theory of argument realization. In Tracy Holloway King & Valeria de Paiva (eds.), *From quirky case to representing space: Papers in honor of Annie Zaenen*, 9–20. Stanford: CSLI Publications.

Alsina, Alex. 1992. On the argument structure of causatives. *Linguistic Inquiry* 23. 517–555.

Alsina, Alex. 1996. *The role of argument structure in grammar: Evidence from Romance* (CSLI Lecture Notes). Stanford: CSLI Publications.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Alsina, Alex. 1997. Causatives in Bantu and Romance. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Complex predicates*, 203–246. Stanford: CSLI Publications.

Alsina, Alex. 1999. Where's the mirror principle? *The Linguistic Review* 16. 1–42. DOI: 10.1515/tlir.1999.16.1.1.

Alsina, Alex. 2001. On the nonsemantic nature of argument structure. *Language Sciences* 23(4–5). 355–389. DOI: 10.1016/s0388-0001(00)00030-9.

Alsina, Alex & Sam A. Mchombo. 1990. The syntax of applicatives in Chicheŵa: Problems for a theta theoretic asymmetry. *Natural Language & Linguistic Theory* 8(4). 493–506. DOI: 10.1007/bf00133691.

Alsina, Alex & Sam A. Mchombo. 1993. Object asymmetries and the Chicheŵa applicative construction. In Sam A. Mchombo (ed.), *Theoretical aspects of Bantu grammar 1*, 17–45. Stanford: CSLI Publications.

Alsina, Alex, K. P. Mohanan & Tara Mohanan. 2005. How to get rid of the comp. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '05 conference*, 21–41. Stanford: CSLI Publications.

Andrews, Avery D. 2007. Projections and glue for clause-union complex predicates. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '07 conference*, 44–65. Stanford: CSLI Publications.

Andrews, Avery D. 2008. The role of pred in LFG+Glue. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '08 conference*, 46–67. Stanford: CSLI Publications.

Andrews, Avery D. & Christopher D. Manning. 1999. *Complex predicates and information spreading in LFG*. Stanford: CSLI Publications.

Asudeh, Ash. 2001. Linking, optionality, and ambiguity in Marathi. In Peter Sells (ed.), *Formal and empirical issues in Optimality Theoretic Syntax*, 257–312. Stanford: CSLI Publications.

Asudeh, Ash. 2012. *The logic of pronominal resumption* (Oxford Studies in Theoretical Linguistics). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199206421.001.0001.

Asudeh, Ash. 2021. The unrealized and the unheard. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 22–44. Oxford: Oxford University Press. DOI: 10.1093/oso/9780192844842.003.0003.

Asudeh, Ash. 2022. Glue semantics. *Annual Review of Linguistics* 8. 321–341. DOI: 10.1146/annurev-linguistics-032521-053835.

Asudeh, Ash. Forthcoming. Glue Semantics. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2013. Constructions with Lexical Integrity. *Journal of Language Modelling* 1(1). 1–54. DOI: 10.15398/jlm.v1i1.56.

Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '12 conference*, 64–84. Stanford: CSLI Publications.

Asudeh, Ash, Gianluca Giorgolo & Ida Toivonen. 2014. Meaning and valency. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 68–88. Stanford: CSLI Publications.

Baker, Mark. 2006 [1983]. Objects, themes and lexical rules in Italian. In Miriam Butt & Tracy Holloway King (eds.), *Lexical semantics in LFG*, 11–57. Stanford: CSLI Publications. Originally published as Levin, Rappaport & Zaenen (1983: 1–45).

Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Belyaev, Oleg. Forthcoming(a). Core concepts of LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Belyaev, Oleg. Forthcoming(b). Grammatical functions: Their classification and properties. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Berman, Judith. 2007. Functional identification of complement clauses in German and the specification of COMP. 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*, 69–83. Stanford: CSLI Publications.

Bögel, Tina, Miriam Butt, Annette Hautli & Sebastian Sulger. 2009. Urdu and the modular architecture of ParGram. In *Proceedings of the Conference on Language and Technology 2009 (CLT09)*, 1–7. Lahore: Center for Research in Urdu Language Processing (CRULP).

Bolinger, Dwight. 1977. *Meaning and form*. London: Longman.

Börjars, Kersti & John J. Lowe. Forthcoming. Noun phrases in LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Börjars, Kersti, Rachel Nordlinger & Louisa Sadler. 2019. *Lexical-Functional Grammar: An introduction*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/9781316756584.

Bresnan, Joan. 1978. A realistic transformational grammar. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 1–59. Cambridge, MA: The MIT Press.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Bresnan, Joan. 1980. Polyadicity: Part I of a theory of lexical rules and representations. In Teun Hoekstra, Harry van der Hulst & Michael Moortgat (eds.), *Lexical grammar*, 97–121. Dordrecht: Foris Publications.

Bresnan, Joan. 1982. The passive in lexical theory. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 3–86. Cambridge, MA: The MIT Press.

Bresnan, Joan. 1990. Monotonicity and the theory of relation-changes in LFG. *Language Research* 26(4). 637–652.

Bresnan, Joan. 1994. Locative inversion and the architecture of Universal Grammar. *Language* 70(1). 72–131. DOI: 10.2307/416741.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, Joan. 2003. Lexical syntax 25 years later: a retrospective and prospective look at the dative alternation in LFG. Handout of paper given at the LFG03 conference, University at Albany, State University of New York. https://web.stanford.edu/~bresnan/slides-lfg03.pdf.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base* (Studies in Generative Grammar), 77–96. Berlin: Mouton de Gruyter.

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, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. KNAW.

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 & Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21. 147–186.

Bresnan, Joan & Annie Zaenen. 1990. Deep unaccusativity in LFG. In Katazyna Dziwirek, Patrick Farrell & Errapel Mejías-Bikandi (eds.), *Grammatical relations: A cross-theoretical perspective*, 45–57. Stanford: CSLI Publications.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu* (Dissertations in Linguistics). Stanford: CSLI Publications.

Butt, Miriam. 2006. *Theories of case*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9781139164696.

Butt, Miriam. 2014. Control vs. complex predication: Identifying non-finite complements. *Natural Language & Linguistic Theory* 32. 165–190. DOI: 10.1007/s11049-013-9217-5.

Butt, Miriam, Tina Bögel, Annette Hautli, Sebastian Sulger & Tafseer Ahmed. 2012. Identifying Urdu complex predication via bigram extraction. In *Proceedings of the 24th international conference on computational linguistics (coling 12)*, 409–424. Mumbai, India: The COLING 2012 Organizing Committee. https://aclanthology.org/C12-1026.

Butt, Miriam, Mary Dalrymple & Anette Frank. 1997. An architecture for linking theory in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '97 conference*, 1–16. Stanford: CSLI Publications.

Butt, Miriam & Tracy Holloway King. 2007. Urdu in a parallel grammar development environment. *Language Resources and Evaluation* 41(2). Special Issue on Asian Language Processing: State of the Art Resources and Processing, 191–207. DOI: 10.1007/s10579-007-9042-8.

Butt, Miriam, Tracy Holloway King & John T. Maxwell III. 2003. Complex predication via restriction. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '03 conference*, 92–104. Stanford: CSLI Publications.

Butt, Miriam, Tracy Holloway King & Gillian Ramchand. 2010. Complex predication: Who made the child pinch the elephant? In Linda Ann Uyechi & Lian-Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 231–256. Stanford: CSLI Publications.

Chisarik, Erika & John R. Payne. 2001. Modelling possessor constructions in LFG: English and Hungarian. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '01 conference*, 33–46. Stanford: CSLI Publications.

Chisarik, Erika & John R. Payne. 2003. Modelling possessor constructions in LFG: English and Hungarian. In Miriam Butt & Tracy Holloway King (eds.), *Nominals: Inside and out*, 181–199. Stanford: CSLI Publications.

Clark, Eve V. 1987. The principle of contrast: A constraint on language acquisition. In Brian MacWhinney (ed.), *Mechanisms of language acquisitions*, 1–33. Lawrence Erlbaum Associates.

Comrie, Bernard. 1974. Causatives and Universal Grammar. *Transactions of the Philological Society* 73(1). 1–32. DOI: 10.1111/j.1467-968X.1974.tb01155.x.

Comrie, Bernard. 1993. Argument structure. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: An international handbook of contemporary research*, vol. 1, 903–914. Berlin: Walter de Gruyter.

Crouch, Richard, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2011. *XLE Documentation*. Xerox Palo Alto Research Center. Palo Alto, CA. https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle_toc.html.

Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge University Press.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Dalrymple, Mary. 1993. *The syntax of anaphoric binding*. Stanford: CSLI Publications.

Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach* (Language, Speech, and Communication). Cambridge, MA: The MIT Press. DOI: 10.7551/mitpress/6169.001.0001.

Dalrymple, Mary. 2001. *Lexical Functional Grammar* (Syntax and Semantics 34). New York: Academic Press. DOI: 10.1163/9781849500104.

Dalrymple, Mary, Angie Hinrichs, John Lamping & Vijay Saraswat. 1993. The resource logic of complex predicate interpretation. In *Proceedings of ROCLING 1993*, 3–21. http://www.aclclp.org.tw/rocling/1993/K01.pdf.

Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '04 conference*, 199–208. Stanford: CSLI Publications.

Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. Stanford: CSLI Publications.

Dalrymple, Mary, John Lamping, Fernando Pereira & Vijay Saraswat. 1999. Overview and introduction. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. Cambridge, MA: The MIT Press.

Dalrymple, Mary, John Lamping & Vijay Saraswat. 1993. LFG semantics via constraints. In *Proceedings of the 6th conference of the European chapter of the ACL (EACL 1993)*, 97–105. Association for Computational Linguistics. DOI: 10.3115/976744.976757.

Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '00 conference*, 104–121. Stanford: CSLI Publications.

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.

Davis, Anthony R. 2011. Thematic roles. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 1, 399–420. Berlin: De Gruyter Mouton.

Davis, Anthony R. & Jean-Pierre Koenig. 2000. Linking as constraints on word classes in a hierarchical lexicon. *Language* 6(1). 59–91. DOI: 10.1353/lan.2000.0068.

Davis, Anthony R. & Jean-Pierre Koenig. 2021. The nature and role of the lexicon in HPSG. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook.* Berlin: Language Science Press.

Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67. 547–619. DOI: 10.2307/415037.

Dryer, Matthew S. 1986. Primary objects, secondary objects, and antidative. *Language* 62(4). 808–845. DOI: 10.2307/415173.

Eckert, Penelope. 2018. *Meaning and linguistic variation: The third wave in sociolinguistics.* Cambridge University Press. DOI: 10.1017/9781316403242.

Falk, Yehuda N. 2001a. Constituent structure and grammatical functions in the Hebrew action nominal. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '01 conference*, 83–103. Stanford: CSLI Publications.

Falk, Yehuda N. 2001b. *Lexical-Functional Grammar: An introduction to parallel constraint-based syntax.* Stanford: CSLI Publications.

Fillmore, Charles J. 1968. The case for case. In Emmon W. Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 1–88. New York: Holt, Rinehart & Winston.

Fillmore, Charles J. 1970. The grammar of hitting and breaking. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 120–133. Waltham, MA: Ginn.

Findlay, Jamie Y. 2016. Mapping theory without argument structure. *Journal of Language Modelling* 4(2). 293–338. DOI: 10.15398/jlm.v4i2.171.

Findlay, Jamie Y. 2020. Mapping Theory and the anatomy of a verbal lexical entry. In Miriam Butt & Ida Toivonen (eds.), *Proceedings of the LFG '20 conference*, 127–147. Stanford: CSLI Publications.

Findlay, Jamie Y. 2021. Meaning in LFG. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 340–374. Oxford: Oxford University Press. DOI: 10.1093/oso/9780192844842.003.0020.

Forst, Martin. 2006. comp in (parallel) grammar writing. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '06 conference*, 222–239. Stanford: CSLI Publications.

Gary, Judith Olmsted & Edward L. Keenan. 1977. On collapsing grammatical relations in universal grammar. In Peter Cole & Jerrold M. Sadock (eds.), *Grammatical relations* (Syntax and Semantics 8), 83–120. New York: Academic Press.

Gawron, Jean Mark. 1983. *Lexical representations and the semantics of complementation.* University of California, Berkeley. (Doctoral dissertation).

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.

Grimshaw, Jane. 1988. Adjuncts and argument structure. *Occasional Paper* 36.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: The MIT Press.

Halvorsen, Per-Kristian. 1983. Semantics for Lexical-Functional Grammar. *Linguistic Inquiry* 14. 567–615.

Harley, Heidi. 2011. A Minimalist approach to argument structure. In Cedric Boeckx (ed.), *The Oxford handbook of linguistic Minimalism*, 427–448. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199549368.013.0019.

Hemmings, Charlotte. 2012. *Causatives and applicatives: The case of polysemy in Javanese*. London: School of Oriental & African Studies. (M.Sc. thesis).

Her, One-Soon. 2003. Chinese inversion constructions within a simplified LMT. *Journal of Chinese Linguistics Monograph Series* (19). 1–31.

Her, One-Soon. 2013. Lexical mapping theory revisited. In Tracy Holloway King & Valeria de Paiva (eds.), *From quirky case to representing space: Papers in honor of Annie Zaenen*, 47–60. Stanford: CSLI Publications.

Homola, Petr & Matt Coler. 2013. Causatives as complex predicates without the restriction operator. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '13 conference*, 316–334. Stanford: CSLI Publications.

Ishikawa, Akira. 1985. *Complex predicates and lexical operations in Japanese*. Stanford: Stanford University. (Doctoral dissertation).

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: The MIT Press.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: The MIT Press.

Joshi, Smita. 1993. *Selection of grammatical and logical functions in Marathi*. Stanford: Stanford University. (Doctoral dissertation).

Kaplan, Ronald M. 1989. The formal architecture of Lexical-Functional Grammar. *Journal of Information Science and Engineering* 5. 305–322. Revised version published in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 7–27).

Kaplan, Ronald M. 1995. The formal architecture of Lexical-Functional Grammar. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.), *Formal issues in Lexical-Functional Grammar*, 7–27. Stanford: CSLI Publications. Earlier version published as Kaplan (1989).

Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. Cambridge, MA: The MIT Press. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 29–130).

Kaplan, Ronald M. & Paula S. Newman. 1997. Lexical resource reconciliation in the Xerox Linguistic Environment. In *Proceedings of the ACL workshop on Computational Environments for Grammar Development and Engineering*. Association for Computational Linguistics.

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.

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.

Kelling, Carmen. 2003. French psych verbs and derived nouns. In Miriam Butt & Tracy Holloway King (eds.), *Nominals inside and out*, 151–180. Stanford: CSLI Publications.

Kibort, Anna. 2001. The Polish passive and impersonal in Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '01 conference*, 163–183. Stanford: CSLI Publications.

Kibort, Anna. 2004. *Passive and passive-like constructions in English and Polish*. Cambridge, UK: University of Cambridge. (Doctoral dissertation).

Kibort, Anna. 2006. On three different types of subjectlessness and how to model them in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '06 conference*, 289–309. Stanford: CSLI Publications.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '07 conference*, 250–270. Stanford: CSLI Publications.

Kibort, Anna. 2008. On the syntax of ditransitive constructions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '08 conference*, 312–332. Stanford: CSLI Publications.

Kibort, Anna. 2009. Intermediary agents and unexpressed pronouns. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '09 conference*, 378–398. Stanford: CSLI Publications.

Kibort, Anna. 2013. Objects and lexical mapping theory. Presented at the 18th International Lexical Functional Grammar Conference, University of Debrecen. Extended abstract available at https://web.stanford.edu/group/cslipublications/cslipublications/LFG/18/abstracts/lfg13abs-kibort.html.

Kibort, Anna. 2014. Mapping out a construction inventory with (Lexical) Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 262–282. Stanford: CSLI Publications.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Kibort, Anna & Joan Maling. 2015. Modelling the syntactic ambiguity of the active vs. passive impersonal in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '15 conference*, 145–165. Stanford: CSLI Publications.

Kisseberth, Charles W. & M. I. Abasheikh. 1977. The object relationship in Chi-Mwi:ni, a Bantu language. In Peter Cole & Jerrold M. Sadock (eds.), *Grammatical relations* (Syntax and Semantics 8). New York: Academic Press.

Kuhn, Jonas. 2001. Resource sensitivity in the syntax-semantics interface: Evidence from the German split NP construction. In W. Detmar Meurers & Tibor Kiss (eds.), *Constraint-based approaches to Germanic syntax* (Studies in Constraint-Based Lexicalism), 177–216. Stanford: CSLI Publications.

Laczkó, Tibor. 2000. Derived nominals, possessors, and lexical mapping theory in Hungarian DPs. In Miriam Butt & Tracy Holloway King (eds.), *Argument realization*, 189–227. Stanford: CSLI Publications.

Laczkó, Tibor. 2003. On oblique arguments and adjuncts in Hungarian event nominals. In Miriam Butt & Tracy Holloway King (eds.), *Nominals: Inside and out*, 201–234. Stanford: CSLI Publications.

Laczkó, Tibor. 2004. Grammatical functions, LMT, and control in the Hungarian DP revisited. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '04 conference*, 313–333. Stanford: CSLI Publications.

Laczkó, Tibor. 2007. Revisiting possessors in Hungarian DPs: A new perspective. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '07 conference*, 343–362. Stanford: CSLI Publications.

Laczkó, Tibor. 2010. A new account of possessors and event nominals in Hungarian. In Artemis Alexiadou & Monika Rathert (eds.), *The semantics of nominalizations across languages and frameworks*, 81–106. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110226546.83.

Levin, Beth. 1993. *English verb classes and alternations*. Chicago: University of Chicago Press.

Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization* (Research Surveys in Linguistics). Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9780511610479.

Levin, Lori S. 1986. *Operations on lexical forms: Unaccusative rules in Germanic languages*. Cambridge, MA: Massachussetts Institute of Technology. (Doctoral dissertation).

Levin, Lori S., Malka Rappaport & Annie Zaenen (eds.). 1983. *Papers in Lexical-Functional Grammar*. Bloomington, IN: Indiana University Linguistics Club.

Lødrup, Helge. 2012. In search of a nominal COMP. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '12 conference*, 383–404. Stanford: CSLI Publications.

Lovestrand, Joseph. 2018. *Serial verb constructions in Barayin: Typology, description and Lexical Functional Grammar*. Oxford: University of Oxford. (D.Phil. Thesis). https://ora.ox.ac.uk/objects/uuid:39406562-02d3-46f5-abf3-180d22225925.

Lovestrand, Joseph. 2020. F-structure and s-structure of Urdu complex predicates. In Miriam Butt & Ida Toivonen (eds.), *Proceedings of the LFG '20 conference*, 232–249. Stanford: CSLI Publications.

Lowe, John J. 2014. Gluing meanings and semantic structures. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 387–407. Stanford: CSLI Publications.

Lowe, John J. 2015. Degrees of degrammaticalization: A lexical sharing approach to the English possessive. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '15 conference*, 208–228. Stanford: CSLI Publications.

Lowe, John J. 2016. Complex predicates: An LFG+glue analysis. *Journal of Language Modelling* 3. 413–462. DOI: 10.15398/jlm.v3i2.125.

Lowe, John J. 2017. *Transitive nouns and adjectives: Evidence from Early Indo-Aryan.* Oxford: Oxford University Press.

Lowe, John J., Adriana Molina-Muñoz & Antonia Ruppel. 2021. Raising and passive in Sanskrit. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG '21 conference*, 243–263. https://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG-2021/lfg2021-lmr.pdf.

Manning, Christopher D. & Ivan A. Sag. 1999. Dissociations between argument structure and grammatical relations. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and constructional aspects of linguistic explanation*, 63–77. Stanford: Stanford University Press.

Markantonatou, Stella. 1995. Modern Greek deverbal nominals: An LMT approach. *Journal of Linguistics* 31(2). 55–58. DOI: 10.1017/s0022226700015619.

Matsumoto, Yo. 1992. *On the wordhood of complex predicates in Japanese.* Stanford: Stanford University. (Doctoral dissertation).

Matsumoto, Yo. 1996. *Complex predicates in Japanese: A syntactic and semantic study of the notion 'word'.* Stanford: CSLI Publications.

Mohanan, Tara. 1990. *Arguments in Hindi.* Stanford: Stanford University. (Doctoral dissertation).

Mohanan, Tara. 1994. *Argument structure in Hindi* (Dissertations in Linguistics). Stanford: CSLI Publications.

Newmeyer, Frederick J. 2002. Optimality and functionality: A critique of functionally-based optimality-theoretic syntax. *Natural Language & Linguistic Theory* 20(1). 43–80. DOI: 10.1023/A:1014290005775.

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Parsons, Terence. 1990. *Events in the semantics of English: A study in subatomic semantics* (Current Studies in Linguistics 19). Cambridge, MA: The MIT Press.

Patejuk, Agnieszka & Adam Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the joint 2016 conference on Head-Driven Phrase Structure Grammar and Lexical Functional Grammar*, 541–559. Stanford: CSLI Publications.

Perlmutter, David M. 1978. Impersonal passives and the unaccusative hypothesis. In Jeri J. Jaeger, Anthony C. Woodbury, Farrell Ackerman, et al. (eds.), *Proceedings of the 4th annual meeting of the Berkeley Linguistics Society*, 157–189. Berkeley: Berkeley Linguistics Society. DOI: 10.3765/bls.v4i0.2198.

Perlmutter, David M. & Paul M. Postal. 1977. Toward a universal characterization of passivization. In Kenneth Whistler, Robert Van Valin Jr., Chris Chiarello, Jeri J. Jaeger, Miriam Petruck, Henry Thompson, Ronya Javkin & Anthony Woodbury (eds.), *Proceedings of the 3rd annual meeting of the Berkeley Linguistics Society*, 394–417. Berkeley: Berkeley Linguistics Society. DOI: 10.3765/bls.v3i0.2266.

Peters, Stanley & R. W. Ritchie. 1973. On the generative power of transformational grammars. *Information Sciences* 6. 49–83. DOI: 10.1016/0020-0255(73)90027-3.

Pollard, Carl. 1997. The nature of constraint-based grammar. *Linguistic Research* 15. 1–18. http://isli.khu.ac.kr/journal/content/data/15/1.pdf.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press & CSLI Publications.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in Generative Grammar*. Tech. rep. 2. New Brunswick, NJ: Rutgers University Center for Cognitive Science. Published as Prince & Smolensky (2004).

Prince, Alan & Paul Smolensky. 2004. *Optimality theory: Constraint interaction in Generative Grammar*. Oxford/Malden: Blackwell. DOI: 10.1002/9780470759400.

Przepiórkowski, Adam. 2016. How *not* to distinguish arguments from adjuncts in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the joint 2016 conference on Head-Driven Phrase Structure Grammar and Lexical Functional Grammar*, 560–580. Stanford: CSLI Publications.

Przepiórkowski, Adam. 2017. A full-fledged hierarchical lexicon in LFG: The FrameNet approach. In Victoria Rosén & Koenraad De Smedt (eds.), *The very model of a modern linguist – In honor of Helge Dyvik*, 202–219. Bergen: Bergen Language & Linguistics Studies (BeLLS). DOI: 10.15845/bells.v8i1.1336.

Ramchand, Gillian. 1997. *Aspect and predication*. Oxford: Oxford University Press.

Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9780511486319.

Ramchand, Gillian. 2014. Argument structure and argument structure alternations. In Marcel den Dikken (ed.), *The Cambridge handbook of generative syntax*, 265–321. Cambridge, UK: Cambridge University Press.

Rappaport, Malka. 2006 [1983]. On the nature of derived nominals. In Miriam Butt & Tracy Holloway King (eds.), *Lexical semantics in LFG*, 121–148. Stanford: CSLI Publications. Originally published as Levin, Rappaport & Zaenen (1983: 113–142).

Rappaport Hovav, Malka & Beth Levin. 2007. Deconstructing thematic hierarchies. 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*, 385–402. Stanford: CSLI Publications.

Sadler, Louisa & Andrew Spencer. 1998. Morphology and argument structure. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 206–236. Oxford: Blackwell. DOI: 10.1002/9781405166348.ch10.

Saiki, Mariko. 1987. *On the manifestations of grammatical functions in the syntax of Japanese nominals*. Stanford: Stanford University. (Doctoral dissertation).

Schätzle, Christin. 2018. *Dative subjects: Historical change visualized*. Konstanz: University of Konstanz. (Doctoral dissertation).

Simpson, Jane. 2006 [1983]. Resultatives. In Miriam Butt & Tracy Holloway King (eds.), *Lexical semantics in LFG*, 149–161. Stanford: CSLI Publications. Originally published as Levin, Rappaport & Zaenen (1983: 143–157).

Speas, Margaret. 1990. *Phrase structure in natural language*. Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-009-2045-3.

Sulger, Sebastian. 2013. Nominal argument structure and the stage/indiviual-level contrast in Hindi-Urdu. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '12 conference*, 582–602. Stanford: CSLI Publications.

Szűcs, Péter. 2018. A comp-less approach to Hungarian complement clauses. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '18 conference*, 325–342. Stanford: CSLI Publications.

Talmy, Leonard. 1978. Figure and ground in complex sentences. In Joseph H. Greenberg (ed.), *Universals of human language*, vol. 4, 625–649. Stanford: Stanford University Press.

Taylor, Roxanne. 2023. *Argument realisation and argument structure in the Old English eventive noun phrase*. University of Manchester. (Doctoral dissertation).

Jamie Y. Findlay, Roxanne Taylor & Anna Kibort

Toivonen, Ida. 2013. English benefactive NPs. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '13 conference*, 503–523. Stanford: CSLI Publications.

Williams, Alexander. 2015. *Arguments in syntax and semantics* (Key Topics in Syntax). Cambridge, UK: Cambridge University Press. DOI: 10 . 1017 / cbo9781139042864.

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 & Elisabet Engdahl. 1994. Descriptive and theoretical syntax in the lexicon. In Beryl T. S. Atkins & Antonio Zampolli (eds.), *Computational approaches to the lexicon*, 181–212. Oxford: Clarendon Press.

Chapter 3

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 character-istics in many typological and descriptive works. Their properties provide fertile grounds for testing assumptions in syntactic theories. In this chapter, we demon-strate that the parallel correspondence architecture of LFG can be used as a pow-erful tool for language-specific linguistic analysis, while also precisely capturing the cross-linguistic differences within and between Western and Eastern AN lan-guages. LFG is flexible in incorporating analytical tests, such as adverbial insertion and clitic placement for examining constituency; reflexive binding, nominal mark-ing 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 lan-guage 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 east-ern point. In fact, this geographical spread is a historical outcome of the prehis-toric 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 group[1] 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

---

[1]In 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 group[2] 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 forthcoming [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

---

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

effectively demonstrated from the LFG perspective where, crucially, no movement is required (cf. Sells's (forthcoming [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 4.1), and the use of case-marking flagging and applicative constructions (cf. Section 4.2). 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 4.3).

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 Kambera in Sumba (Klamer 1998) and Wooi 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 theoret-

ical significance, such as ergativity or the complex interconnection between surface 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 forthcoming(b) [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, while grammatical functions (GFs) of arguments and grammatical features are dealt with in f-structure, as detailed in Belyaev forthcoming(a) [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 & Kibort forthcoming [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)[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]$_{SUBJ}$ ∅-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. Tr<em>akaw [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 re-
lated argument realisations. First, verbal voice morphology marks SUBJ selection
(cf. Section 4). The ACTOR VOICE (AV), indicated by *ng-* in Balinese (1)a and ⟨*em*⟩ in
Puyuma (2)a, 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 (1)b, whereas the CONVEYANCE VOICE
(CV) in Puyuma selects a peripheral role, as in the beneficiary 'mother' in (2)b[5]

---

[3]In 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 genealogi-
cal and typological relations between languages. WMP and CEMP stand for Western Malayo-
Polynesian and Central-Eastern Malayo-Polynesian respectively.

[4]Following 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 argu-
ment 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).

[5]In 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).

(see Section 4.1 for the properties of SUBJ/PIVOT and the typology of AN voice systems). The voices in these AN languages are called SYMMETRICAL VOICE. Evidence for their symmetricality 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, symmetricality 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 (1)b and (2)b 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 (1)b appears immediately after the verb as a bare Noun Phrase (NP), like the OBJ of the AV verb in (1)a. 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 (1)b and CV in Puyuma (2)b, 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 promi-

---

[6]Balinese 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.

[7]Note 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 4.3.
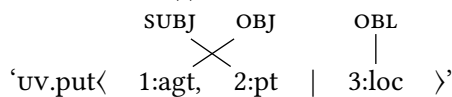
nence, which includes thematic ranking. Although GRs and GFs belong to different layers of structure, the two are interrelated and are mapped together through linking principles. In AN languages, the mapping is marked by verbal voice marking and/or indexing, depending on the available coding resources. 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 (3)a for the Balinese verb *jang* 'put', and its simplified representation is given in (3)b. 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⟩'

---

[8]Syntactic 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.
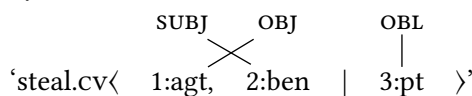
The following should be noted regarding the representation of GRs in (3). First, GRs conceptually reflect event construal and participant roles, which signifies 'who does what to whom'. This is the semantic-conceptual basis underlying valency and transitivity information in the syntacticised a-structure. 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 (1)b and Puyuma CV in (2)b 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 (1)b

                   SUBJ   OBJ   OBL

           'UV.put⟨ 1:agt, 2:pt | 3:loc ⟩'

    b. Puyuma CV in (2)b

                   SUBJ   OBJ   OBL

      'steal.cv⟨ 1:agt, 2:ben | 3:pt ⟩'

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 ngenehang [awakne]_OBJ.
        3   AV.think   self.3
      '(S)he thought of herself/himself.'

                            SUBJ   OBJ

              'AV.think⟨1:agt.'3'$_i$,   2:th.'self.3'$_i$ ⟩'

I Wayan Arka & Li-Chen Yeh
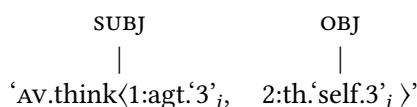
  b. [Awakne]$_{SUBJ}$ kenehang[=a]$_{OBJ}$.
      self.3   uv.think=3
      '(S)he thought of *herself/himself.*'

                                              SUBJ   OBJ

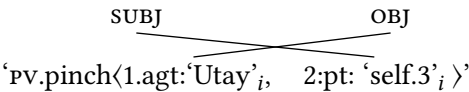   'uv.think⟨1:agt.'3'$_i$,   2:th.'self.3'$_i$⟩'
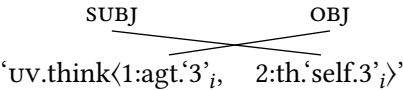
(6)   Kavalan (Formosan) (Shen 2005: 57)

  a. K⟨em⟩nit [ci Utay]$_{SUBJ}$ [tu izipna]OBL.
      ⟨AV⟩pinch PN Utay   OBL body:3GEN
      'Utay pinched at himself.'

                                          SUBJ   OBL
                                            | |
                        'AV.pinch⟨1:agt:'Utay'$_i$, | 2:pt: 'self.3'$_i$ ⟩'
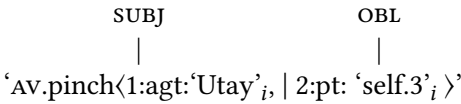
  b. Kenit-an=na   [ni Utay]$_{OBJ}$ [ya izipna]$_{SUBJ}$.
      pinch-PV=3ERG ERG Utay   ABS body:3GEN
      'Utay pinched *himself.*'

                                          SUBJ   OBJ

                        'PV.pinch⟨1.agt:'Utay'$_i$,   2:pt: 'self.3'$_i$ ⟩'

  c. f-structure of sentence (6)a

$$
\begin{bmatrix}
\text{PRED} & \text{'PINCH⟨SUBJ OBL⟩'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'UTAY'} \\ \text{NTYPE} & \text{PROPER} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \text{'3'} \\ \text{NUM} & \text{SG} \end{bmatrix}_i \end{bmatrix} \\
\text{OBL} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{REFLEXIVE} \\ \text{INDEX} & i \end{bmatrix} \\
\text{VOICE-TYPE} & \text{AV}
\end{bmatrix}
$$

  d. f-structure of sentence (6)b

$$
\begin{bmatrix}
\text{PRED} & \text{'PINCH⟨SUBJ OBJ⟩'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{REFLEXIVE} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'UTAY'} \\ \text{NTYPE} & \text{PROPER} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \text{'3'} \\ \text{NUM} & \text{SG} \end{bmatrix}_i \end{bmatrix} \\
\text{VOICE-TYPE} & \text{PV}
\end{bmatrix}
$$

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 (5)a and (5)b share the same a-structure, but differ in their respective mapping to GFs: consider the crossing line in (5)b in the UV structure. In Kavalan, both (6)a and (6)b are similar in their a-structure representations except that the non-actor argument in (6)a 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 (5)a and SUBJ in UV in (5)b 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 (6)a (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 (6)b. 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 forthcoming [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 (5)b and (6)b 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]

---

[9]For the sake of brevity, a detailed comparation with other theories is omitted here. We con-

I Wayan Arka & Li-Chen Yeh

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 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

---

fine our illustration to the use of LFG in analysing AN languages. An in-depth comparative discussion of LFG and other frameworks are provided in Part VII.

[10]Some Formosan languages (e.g. Puyuma and Kavalan) have pronominal indexes on verbs. While they closely interact with its 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.

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 forthcoming [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 (7)a, *hira'* 'yesterday' cannot intervene between a transitive verb and its object. The c-structure of (7)a is represented in (7)b. Note that LFG adopts a version of X-bar syntax that allows nonbinary branching, as seen in the top/root node of IP in (7)b.

(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.'

---

[11]While 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 forthcoming [this volume].

I Wayan Arka & Li-Chen Yeh

b.
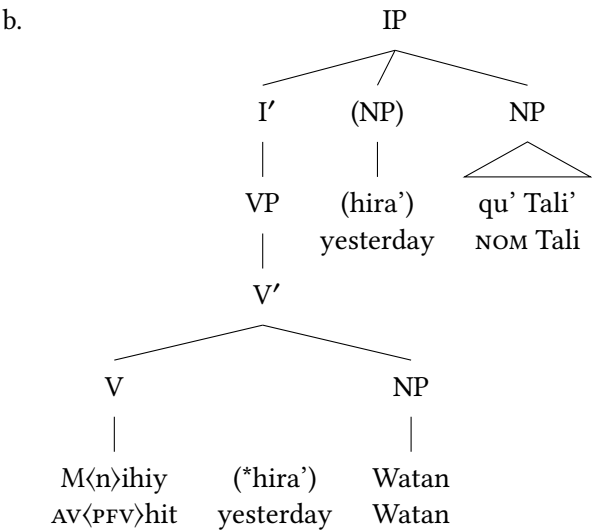
Turning to Tagalog, we can posit that the finite sentence (IP) in this language contains a non-configurational (i.e. exocentric, flat) Sentence (S), as shown in (8)b for the example in (8)a; for further discussion of exocentricity and the category S, see Andrews forthcoming [this volume]. Evidence for 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.'

b.
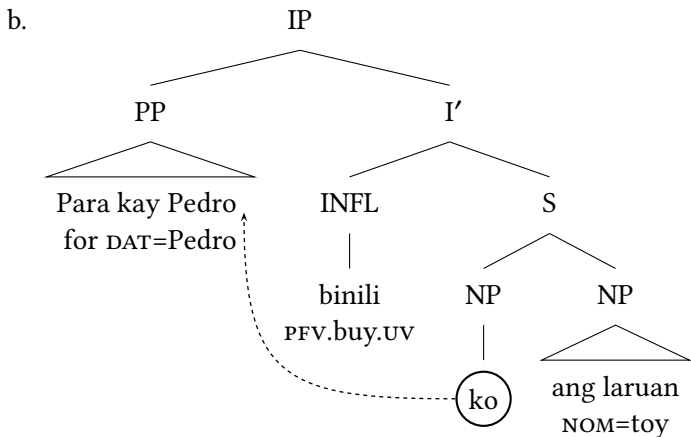
Variation in predicate-initial word order is pragmatically driven, and allows a unit to be 'fronted' to a sentence-initial position before the verb. This position bears a Discourse Function (DF) and is not uniquely associated with a particular grammatical function. In the Tagalog example (8)a, the fronted DF, the OBL 'for Pedro', structurally occupies the Specifier position of the finite sentence, [Spec,IP], as shown in the c-structure in (8)b. 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.'

---

[12]Kroeger'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 2004). See Halpern (1995), Halpern & Zwicky (1996), King (2005), and Bögel et al. (2010), among others, for further discussion of (2P) clitic placement.

I Wayan Arka & Li-Chen Yeh

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 (10)a-b.

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

    a. [[Ma-desek]$_{\text{V:FOCUS-C}}$ [=siw]$_{\text{SUBJ}}$]$_{\text{IP}}$.
        STAT-belch   =2SG.ABS
     'You *belch*!'

    b. [[M⟨in⟩e-ken]$_{\text{V:FOCUS-C}}$ [=siw]$_{\text{SUBJ}}$ sumay=lia]$_{\text{IP}}$?
        AV ⟨PFV⟩eat   =2SG.ABS rice(meal)=MODAL
     'Have you *eaten* meals?'

    c. [M⟨in⟩e-ken asai paj=   [isiw]$_{\text{SUBJ}}$]$_{\text{IP}}$?
        AV ⟨PFV⟩eat what MODAL= 2SG
     '*What* have you eaten?'

In (10)a-b, the 2P clitic pronoun *=siw* appears as the sole argument of a simple stative intransitive verb (*madesek* 'belch' in (10)a) and an intransitive clause[13] (*meken* 'eat' in actor voice in (10)b). These sentences come with an emphatic focus on the predicates[14] (indicated by italicisation in the free translation; cf. (10)b and (10)c). The free pronoun *isiw* that encodes the SUBJ of a wh-question in sentence (10)c differs from the 2P clitic pronoun in its pragmatic function. Unlike the predicate host in (10)b, there is no emphatic focus on the verb in (10)c, 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,

---

[13]The issue of semantic versus syntactic transitivity of actor-voice clauses in some AN languages is discussed in Section 4.3.

[14]The term "focus" is used in this chapter to refer to the notion in information structure (Zaenen forthcoming [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."

while neutral in case, bears DF for contrastive meaning to encode a highly topical 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 (11)a and VSO with a salient referent of OBJ in (11)b.

(11) Pazeh-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*.'

    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.'

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 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.

---

[15]The 1st person singular pronominal form in Pazeh-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.

(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 (15)a is a pragmatically marked VOS sentence (as seen from its translation).[16] A postverbal object is unacceptable, as depicted in (15)b.

(15) Balinese

a. [[Nunas kopi-ne niki]$_{VP}$, [[tiang]$_{SUBJ}$]$_{IP}$]$_{IP}$.
AV.take coffee-DEF this (pause) 1

'Taking this coffee was what I did.'

b. *[[Nunas]$_{V:FOCUS-C}$, [[tiang]$_{SUBJ}$ [kopi-ne niki]$_{VP}$]$_{IP}$]$_{IP}$.

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 forthcoming [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 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

---

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

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 (16)a (cf. Arka 2011: 78-80 and Arka (2021) for details). The c-structure tree is given in (16)b, showing that the pronominal relativiser *yang* is grammatically OBJ, and the RC is structurally OSV. The f-structure is shown in (16)c.[17]

(16) Colloquial Indonesian[18] (WMP, Indonesia)

 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.'

---

[17]Note 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.

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

I Wayan Arka & Li-Chen Yeh

b.

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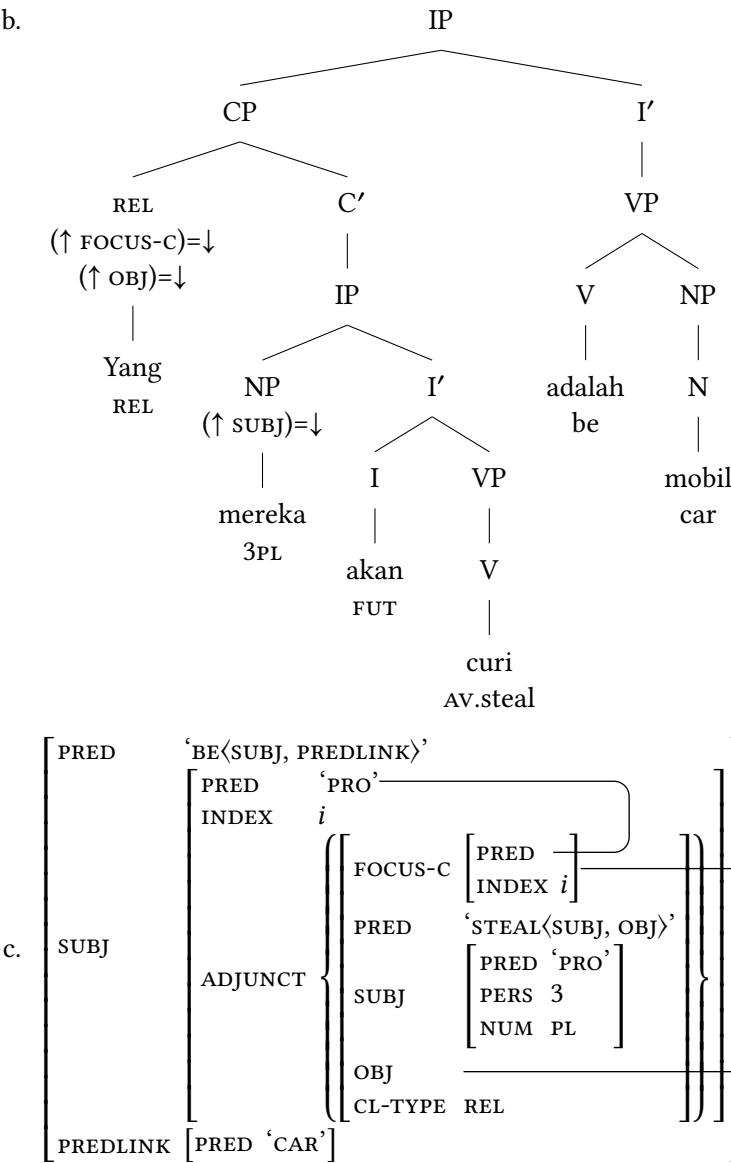
Unlike Indonesian, Balinese has no complementiser C equivalent to English *that* (Arka 2003a).[19] A fronted element bearing a contrastive DF can be analysed as being left-adjoined to IP. This structure is shown by the IP subscripts in example

---

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

(15)a. 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 (17)a. In (17)b, 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 4.3 for further discussion of definiteness). For example, the Balinese generic statement about a cow in (18)a must be in SVO; a VOS variant is unacceptable, as in (18)b.

(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:

I Wayan Arka & Li-Chen Yeh

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

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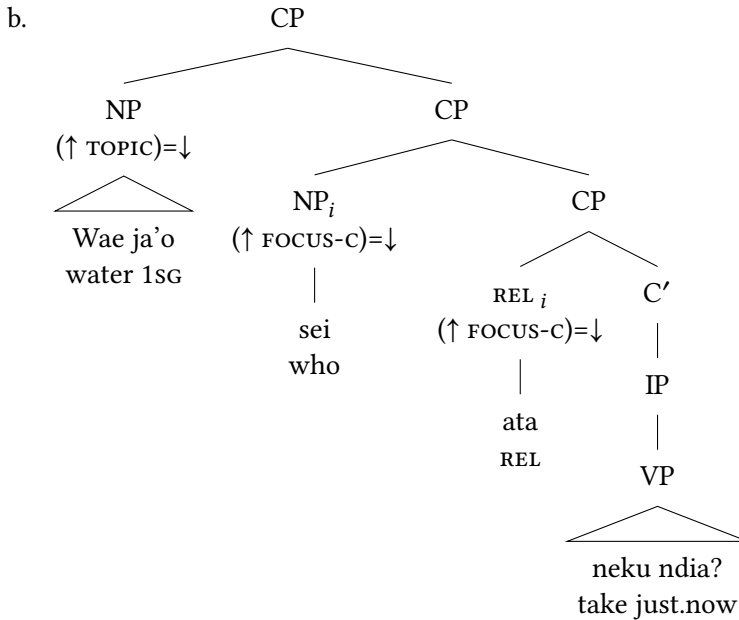
    b.   *Ardi ngedho **nembumai** wolo.

Rongga has developed a relativiser (REL) from the noun meaning 'person', *ata*, exemplified in (20)a. 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 (20)b, which shows that sentence (20)a 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?')

---

[20]It 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).

b.

```
                              CP
                   ┌──────────┴──────────┐
                  NP                     CP
              (↑ TOPIC)=↓        ┌────────┴────────┐
              ┌───┴───┐         NPᵢ                CP
            Wae ja'o        (↑ FOCUS-C)=↓      ┌────┴────┐
            water 1SG            │           REL ᵢ      C′
                               sei      (↑ FOCUS-C)=↓    │
                               who            │          IP
                                             ata         │
                                             REL         VP
                                                      ┌───┴───┐
                                                   neku ndia?
                                                   take just.now
```

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 (21)a 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

---

[21]The 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

I Wayan Arka & Li-Chen Yeh

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).

4 Grammatical functions and alternative argument realisations

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

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 Walpiri. 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).

(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 & Kibort forthcoming [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, Wooi, 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 (21)b. That is, sentence (21)b 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 4.3. We begin by clarifying the subtle and crucial difference between sᴜʙᴊ and ᴘɪᴠᴏᴛ.

4.1 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 forthcoming(a) [this volume]). In this book chapter, we keep the standard LFG conception of sᴜʙᴊ(ᴇᴄᴛ) (i.e. in upper case): it is the surface grammatical subject, the most prominent ɢꜰ on the relational ɢꜰ 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 sᴜʙᴊ from a-subject is necessary to account for symmetrical voice alternations and related properties in AN languages (cf. the separation of ɢꜰ-ɢʀ in Section 2), otherwise certain unusual phenomena, such as the binding of sᴜʙᴊ by oʙᴊ (e.g. in Balinese example [5]) cannot be accounted for.

Furthermore, it is essential to distinguish sᴜʙᴊ from ᴘɪᴠᴏᴛ 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 4.1.2). We also want to emphasize our conception of ᴘɪᴠᴏᴛ as schematised in (22)a: ᴘɪᴠᴏᴛ is at the interface of syntax and pragmatics; that is, it shows grammatical properties (i.e. GF-related, typically intersecting with sᴜʙᴊ, though not always) as well as discourse-pragmatic (ᴅꜰ) properties (e.g. ꜰᴏᴄᴜs-c). In this sense, ᴘɪᴠᴏᴛ is an 'overlay' or 'intersection' of ɢꜰs and ᴅꜰs (Arka 2021). ᴘɪᴠᴏᴛ is evident in the formation of bi-clausal structures, such as relativisation and coordination, and other grammatical mono-clausal structures involving marked ᴅꜰs, such as ꜰᴏᴄᴜs-c in Balinese fronted questions. Informally, licensing sᴜʙᴊ to bear ꜰᴏᴄᴜs-c as ᴘɪᴠᴏᴛ can be represented as (22)b, thus [ᴅꜰ-c=sᴜʙᴊ] ᴘɪᴠᴏᴛ for sᴜʙᴊ ᴘɪᴠᴏᴛ.

(22)   a.

---

[22]The 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).

(Discourse function)   (Grammatical function)

b. [FOCUS-C=SUBJ]$_{PIVOT}$
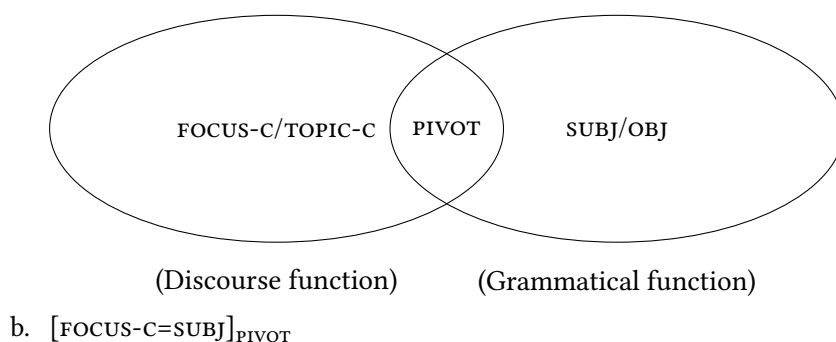
The notion of 'pivot' has been discussed and used in previous typological research. In what follows, we briefly provide some context for the conception of PIVOT adopted in this chapter. Its usage here is broadly aligned to descriptive functional and typological linguistics in place of 'Subject' and 'Topic' 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]

---

[23]Falk'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 forthcoming [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.

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 $\widehat{\text{GF}}$, and their related mapping. The notation of $\widehat{\text{GF}}$ (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 $\widehat{\text{GF}}$ 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 $\widehat{\text{GF}}$ (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 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 (1)a-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 (22)a. 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).

4.1.1 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

    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$_{\text{PATIENT}}$
                          'AV.buy⟨1:agt, 2:pt ⟩'

I Wayan Arka & Li-Chen Yeh

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

<div align="right">

OBJ_{AGENT} SUBJ
'PV.buy⟨ 1:agt,   2:pt ⟩'
</div>

    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.'

<div align="right">

OBJ_{AGENT} OBJ_{PATIENT} SUBJ
'LV.buy⟨ 1:agt,   2:pt,   3:loc ⟩'
</div>

(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.'

<div align="right">

SUBJ OBJ
'AV.spooned.up⟨1:agt,2:pt⟩'
</div>

    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.'

<div align="right">

OBJ SUBJ
'PV.spooned.up⟨1:agt, 2:pt⟩'
</div>

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, 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

---

[24]The 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)

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 sibilation 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 4.2.

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 (23)a-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 (23)b, ⟨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 (25)b-(25)c below, which is a near equivalent of the LV in Tagalog, as seen previously in (23)c:

(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⟩'

---

[25]We thank an anonymous reviewer for pointing this out.

I Wayan Arka & Li-Chen Yeh

b. Ia meli-**nin** **dagang-e ento** baas.
3 AV.buy-APPL trader-DEF that rice
'(S)he bought rice from the trader.'

$$\text{'AV.buy}\langle\text{1:agt, 2:loc/source, 3:th}\quad\rangle\text{'}$$

with columns SUBJ, OBJ, OBJ_THEME above.

c. **Anak-e** **nto** belin-**in** tiang potlot.
person-DEF that UV.buy-APPL 1 pencil
'I bought a pencil from the person.'

$$\text{'UV.buy}\langle\text{1:agt, 2:loc/source, 3:th}\quad\rangle\text{'}$$

with columns OBJ, SUBJ, OBJ_THEME above.

In Balinese, the two-place transitive verb *beli* 'buy' obligatorily requires the applicative marker *-(n)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 (25)a and the unflagged locative/source argument in (25)b. 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 (25)b-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 (25)c. Similar to Formosan/Philippine languages, core arguments are equally selectable as SUBJ in two-way voice systems (i.e. evidencing the symmetricality 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 (26)b.

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

a. **Asép ng**irim buku ka Enéng.
Asep AV.send book to Eneng
'Asep sent a book to Eneng.'

$$\text{'AV.send}\langle\text{1:agt, 2:pt }|\text{3:go}\rangle\text{'}$$

with columns SUBJ, OBJ, OBL above.

---

[26]The noun phrase flagged by *sig* in (25)a 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.

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.'

<div align="right">

SUBJ OBL OBL
'PASSV.send⟨1:pt | 2:agt, 3:go⟩'

</div>

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 (27)a and transitive SUBJ in (27)b. 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 (27)b). 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.'

4.1.2 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 (28)a 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 (28)b 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 does not involve the overlay function of PIVOT. In the Balinese example (28)a, the Q is not fronted. Even for the fronted Q in example (28)b, 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

    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-sti'-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 (30)c in contrast to (30)b), 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 (29)a forms an open question without the presupposition of contrasting entities in the given context (Wei 2009: 348). In contrast, the fronted Q in (30)a 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 Squliq 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 (30)a-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 (28)b 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 (29)a, and a fronted subj, as in (30)a (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 (30)b. Retaining av morphology on the verb renders the structure with a fronted q ungrammatical, as seen in (30)c. Likewise, in contrast to (29)b, the structure is ungrammatical when the fronted q *cima* is obl marked by *-an*, as in (30)d.

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'

---

[27]This is evident in relativisation in familiar languages, like English, where non-subj can be pivot (i.e. gapped in relativisation).

4.2 Non-subj functions: obj and obl

In LFG, there are three non-subj functions: obj, obj$_\theta$, 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$_\theta$ 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) (Squliq Atayal) and *apa* 'what' in (28) (Balinese). Squliq 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 (32)a and is flagged by *te*, and not indexed on the verb. It appears as subj, which is indexed by the enclitic *=ko*,

I Wayan Arka & Li-Chen Yeh

and is optionally cross-referenced by the NOM NP that is flagged by *na* in (32)b.

(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.' (Lit. '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 (32)a 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 (32)b 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 (33)a, locative/goal in (33)b (with the verb containing the locative applicative, *-e*), and recipient/benefactive in (33)c (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']$_{\text{OBJ}}$ [paket]$_{\text{OBJ}_\theta}$.
   elder.sister AV.send-APPL mother   package

'Big Sister sent Mother a package.'   SUBJ   OBJ OBJ$_{\text{THEME}}$
                                   'AV.send.for⟨1:agt,2:goal. 3:th⟩'

c. Sa'diyah melle-**yagi**   [na'-kana']$_{\text{OBJ}}$ [permen]$_{\text{OBJ}_\theta}$.
   Sa'diyah AV.buy-APPL REDUP-child   candy

'Sa'diyah bought the children candy.'

                                   SUBJ   OBJ OBJ$_{\text{THEME}}$
                                  '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$_\theta$ surfacing differently. Consider, firstly, the PP in (33)a, *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 (33)b. 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$_\theta$. This results in a ditransitive structure of SVOO. Likewise, the same restructuring of a-structure occurs with the benefactive applicative in (33)c.

In LFG, the NP *paket* in (33)b is an instance of OBJ$_{\text{THEME}}$ 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 (34)b in contrast to (34)a. 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 (33)b, or SUBJ in UV in (34)a.[28]

(34)   Madurese (Davies 2010: 284)

    a. Ebu'   e-kerem-e   [paket]$_{\text{OBJ}}$ bi' Embuk.
       mother UV-send-APPL package   by elder.sister

      'Mother was sent a package by Big Sister.'

                                  OBJ   SUBJ OBJ$_{\text{THEME}}$
                                'UV.send⟨1:agt,2:goal, 3:th⟩'

---

[28]The 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).

185

I Wayan Arka & Li-Chen Yeh

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 (35)a, but *ngga* in (35)b, since it is thematically beneficiary. Note that the displaced theme, OBJ_{THEME}, in (35)b is DAT. In LFG, the Kambera ditransitive sentence in (35)b has the same a-/f-structures as the Madurese examples in (33)b-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 (35)c 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 forthcoming(b) [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   (↑ OBJ PRED)='PRO'
                           (↑ OBJ PERS)=1
                           (↑ OBJ NUM)=SG
                           (↑ OBJ CASE)=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 (36)a, whereas indefinite OBJ does not, as in (36)b. In an LFG analysis, Palauan DOM can be captured by annotating the suffix slot in the verb formation rule with the constraining equation: (↑ OBJ DEF)=$_c$+. The suffix *-ii* also carries a definiteness feature in its lexical entry, (↑ 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).'

Obliques 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 (33)a, 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 (37)a, and *dra* for an indefinite common noun as in (37)b-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. Tr⟨em⟩aka-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)

4.3 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 (37)a, and *-aw* for PV, patient voice, in (37)c. The AV verbs are also morphologically marked by *-em-* as in (37)b.[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.

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

---

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

   a. T⟨em⟩engedr   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 (23)a. 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 (39)a, the reverse does not hold since an oblique-marked patient can have a definite reading, as seen in (39)b (cf. DOM in Puyuma in Section 4.2). This suggests that in many Philippine-type languages, the coreness 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   ʔucia.
      PFV=drink⟨AV⟩=COMPL 3SG.NOM OBL.CN tea
      'He has drunk the tea.'

---

[30]The 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 straight-forward 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* 'run'. 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 (42)a-b, regardless of its status as the controller (i.e. PV-SUBJ or AV-OBL) in the matrix clause. By contrast, as shown in (42)c, it is not acceptable for the AV patient (i.e. *ci Akian* in (42)c) 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. The status of being controller can vaguely support an accusative analysis of AV constructions.

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

---

[31]However, 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.

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: (↑ CASE)=ABS. The grammar of the language can be globally specified as having a conditional if-then constraint: (↑ SUBJ) ⇒ (↑ 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 (↑ 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 (↑ 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 in (43). DOM across languages commonly draws on different semantic properties. In LFG, these semantic features can be specified together with the semantic

---

[32]The 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 (Himmelmann 2005: 144–149).

case value without affecting the syntactic status of the argument (Butt & King 1991, 2003, Dalrymple & Nikolaeva 2011).
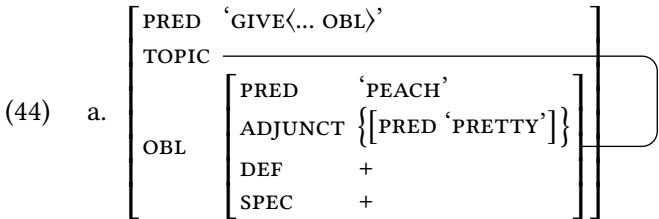
In Pazeh-Kaxabu, DOM encodes the differential information related to the topicality and specificity/definiteness of P (Yeh in preparation). For example, in the AV structure in (43)a, P is realised as OBL due to the ergative system and is flagged by *u* because the referent is definite and topical (reading [i]) or specific indefinite (reading [ii]). The non-specific indefinite P (cf. example (10)b) 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 (43)b. 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 (43)a. 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 (44)a where the reference is definite and specific (cf. Enç 1991; see also Heusinger 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)=+ }

(44) a.

$$
\begin{bmatrix}
\text{PRED} & \text{'GIVE}\langle ... \text{OBL}\rangle\text{'} \\
\text{TOPIC} & \overline{} \\
\text{OBL} & \begin{bmatrix}
\text{PRED} & \text{'PEACH'} \\
\text{ADJUNCT} & \{[\text{PRED 'PRETTY'}]\} \\
\text{DEF} & + \\
\text{SPEC} & +
\end{bmatrix}
\end{bmatrix}
$$

I Wayan Arka & Li-Chen Yeh

$$
\text{b.} \quad \begin{bmatrix} \text{PRED} & \text{`GIVE}\langle \dots \text{OBL}\rangle\text{'} \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{`PEACH'} \\ \text{ADJUNCT} & \{[\text{PRED `PRETTY'}]\} \\ \text{DEF} & - \\ \text{SPEC} & + \end{bmatrix} \end{bmatrix}
$$

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 (45)c and (46)c, respectively. The sigma metavariable ($\uparrow_\sigma$) in the entries ensures the correct mapping or correspondence between semantic a-structure and f-structure, so the constraint represented as ($\uparrow_\sigma$ A) in (45)c 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   ($\uparrow$ PRED)=`PRO'
                   ($\uparrow$ PERS)=3
                   ($\uparrow$ NUM)=SG
                   ($\uparrow$ SUBJ)$_\sigma$ = ($\uparrow_\sigma$ 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 rhët(=**geuh**).

       3   fall(=3)

       'S/he falls.'

    c. *geuh*   CL ($\uparrow$ PRED)= 'PRO'

                         ($\uparrow$ PERS) = 3

                         ($\uparrow$ NUM) = SG

                         ($\uparrow$ {SUBJ|OBJ})$_\sigma$ = ($\uparrow_\sigma$ 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.

I Wayan Arka & Li-Chen Yeh

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 (47)a below, or an XCOMP (without P-flagging) as in (47)b. 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 (47)c, 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 (47)b, but a core argument in (47)c.

(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 luung]_{XCOMP}.
       3 want want-APPL SUBJ AV.build house good
       '(S)he wants build a good house.'

    c. [__ Ngae umah luung]_{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 (48)a 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 (48)b. Note that *coba* 'try' allows different patterns of control, including the so-called double (backward/forward) control given in (48)c. 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).

196

(48) Indonesian

    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 (49)a 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 (49)b, however, the SUBJ is raised and appears as OBL in the matrix clause, flagged by *kana*. Note that raising in (49)b 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=ku   **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.

I Wayan Arka & Li-Chen Yeh

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   wangsi 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-maŋmaŋ yo   [be   aŋ-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 (52)a, as well as XCOMP like in (52)b. 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-lēn   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 across languages, 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 (56)b. 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

I Wayan Arka & Li-Chen Yeh

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 (56)a.

Note that, in the resultative SVC of (56)b, 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 V1 nor V2 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 V2's SUBJ, and the SVC's SUBJ has the same value as the SUBJ of V1. This sharing of values for SUBJ and OBJ is indicated through tags [1] and [2] in (56)a.

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

$$
\begin{bmatrix}
\text{PRED} & \text{`CAUSE.RESULT}\langle\text{SUBJ}[1]: \text{`BLOW}\langle[1]\rangle\text{'}, \text{OBJ}[2]: \text{`FLY.AWAY}\langle[2]\rangle\text{'}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`WIND'} \\ \text{DEF} & + \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`CARDS'} \\ \text{DEF} & + \end{bmatrix}
\end{bmatrix}
$$

    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 (57)a: *(ng)ajak*. The absence of the AV prefix (i.e. *ajak*) gives rise to a comitative reading only, as shown by reading (i) in (57)a; 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 (57)a. The presence of the clausal negator, *tan* 'not', in (57)b 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 (57)b.

(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 (57)a with the bare verb, *ajak*, has one intonational contour (i.e. without a break), while the coordination in (57)b 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 (57)a 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 4.1.2).

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 forthcoming [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 forthcoming [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 (58)a: 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 (58)b below:

(58) Indonesian

    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 (62)a where SUBJ is questioned).

(59) Indonesian
[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)
[Sahaʔ]_{PRED/FOCUS-C} [nu meuliʔ 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

a. [Siapa]_{SUBJ} mem-bunuh perampok itu?
who   AV-kill   robber   that
'Who killed the robber?'

b. Orang itu   membunuh [siapa]_{OBJ}?
person that AV-kill   who
'Who did the person kill?'

c. *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 (62)c above, in contrast to the acceptable in-situ question (62)b. 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 (63)a 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 (63)b, 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 (63)a is identical to that in (63)b. 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]$_{\text{PRED/FOCUS-C}}$ [ata tengo hau]$_{\text{SUBJ}}$?
        who   REL AV.hit you
      'Who hit you? (Lit. 'who is the one hitting you?')' (Semiun 1993: 63)

    b. [Cai]$_{\text{PRED/FOCUS-C}}$ [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 Manggarai, 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]$_{\text{PRED}}$ [ang d⟨um⟩ating]$_{\text{SUBJ}}$?
 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.

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

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

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

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 (67)a 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 (67)a 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 (67)b, where the patient argument is FOCUS-C. Like in Manggarai and Philippine-type languages, the agent in Kambera in (67)b 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$_j$]$_{\text{SUBJ}k}$.
      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 (68)a, the FOCUS-C formation requires no nominalisation as seen in (68)b-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 (68)d. Crucially, the sentences with fronted FOCUS-C NPs in (68)b-d are monoclausal.

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

a. [tau   [na=buno=a sorodadu]$_{\text{RC}}$]$_{\text{NP}}$
    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
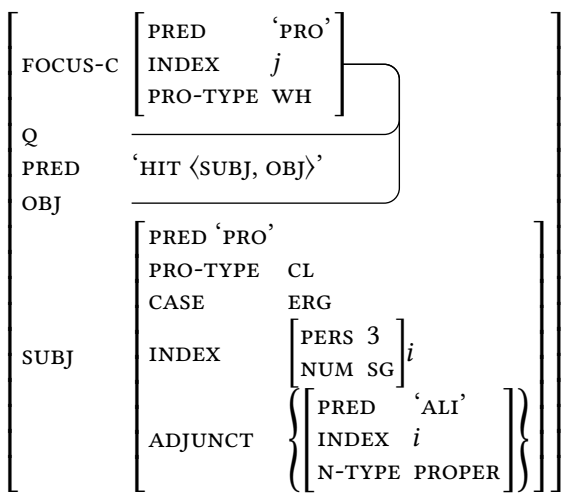
analysis is straightforward for cases involving FOCUS-C with no requirement for nominalisation, as in the Makassar examples of (68)b-d. Here, the sentences are mono-clausal, and the fronted argument is functionally not the head predicate. For the analysis to work, the sentence-initial XP is identified as FOCUS-C and is licensed by the phrase structure rule shown in (69)a below:

(69) a. CP $\longrightarrow$ XP   C′
                              (↑ FOCUS-C)=↓
                              (↑ FOCUS-C)=(↑ GF)

      b. Makassarese Voice Marking:
            i)   AV, *aN-*:   (↑ FOCUS-C) = (↑ SUBJ)$_\sigma$ = (↑$_\sigma$ 1:agent)
            ii)   PV, Clitic$_A$=:   (↑ FOCUS-C) = (↑ OBJ)$_\sigma$ = (↑$_\sigma$ 2:patient)

The two lines of annotation in (69)a impose a sharing between FOCUS-C and any GF, including adjunct. However, there are also other independent language-specific voice selection constraints given in (69)b 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 (69)b.ii, the example in (68)c (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)

I Wayan Arka & Li-Chen Yeh

$$
\begin{bmatrix}
\text{FOCUS-C} & \begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{INDEX} & j \\ \text{PRO-TYPE} & \text{WH} \end{bmatrix} \\
\text{Q} & \\
\text{PRED} & \text{`HIT } \langle \text{SUBJ, OBJ} \rangle\text{'} \\
\text{OBJ} & \\
\text{SUBJ} & \begin{bmatrix} \text{PRED `PRO'} \\ \text{PRO-TYPE} \quad \text{CL} \\ \text{CASE} \quad \text{ERG} \\ \text{INDEX} \quad \begin{bmatrix} \text{PERS 3} \\ \text{NUM SG} \end{bmatrix}_i \\ \text{ADJUNCT} \quad \left\{ \begin{bmatrix} \text{PRED} & \text{`ALI'} \\ \text{INDEX} & i \\ \text{N-TYPE} & \text{PROPER} \end{bmatrix} \right\} \end{bmatrix}
\end{bmatrix}
$$

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 (58)b, 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, as shown in (73) below, shows that the fronted Q is the FOCUS-C PREDLINK.

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

---

[33]The 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).

$$
\begin{bmatrix}
\text{PRED} & \text{`BE⟨SUBJ, PREDLINK⟩'} \\
\text{SUBJ} & \begin{bmatrix}
\begin{matrix} \text{PRED} & [1]\,\text{`PRO'} \\ \text{INDEX} & i \end{matrix} \\
\text{ADJUNCT} \left\{ \begin{bmatrix}
\text{PIVOT} & \begin{bmatrix} \text{PRED} & [1] \\ \text{INDEX} & i \end{bmatrix} \\
\text{PRED} & \text{`KILL⟨SUBJ OBJ⟩'} \\
\text{SUBJ} & \\
\text{OBJ} & \begin{bmatrix} \text{PRED `THE ROBBER'} \end{bmatrix} \\
\text{CL-TYPE} & \text{REL}
\end{bmatrix} \right\}
\end{bmatrix} \\
\text{PREDLINK} & \begin{bmatrix} \text{PRED `JOHN'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{FOCUS-C} &
\end{bmatrix}
$$
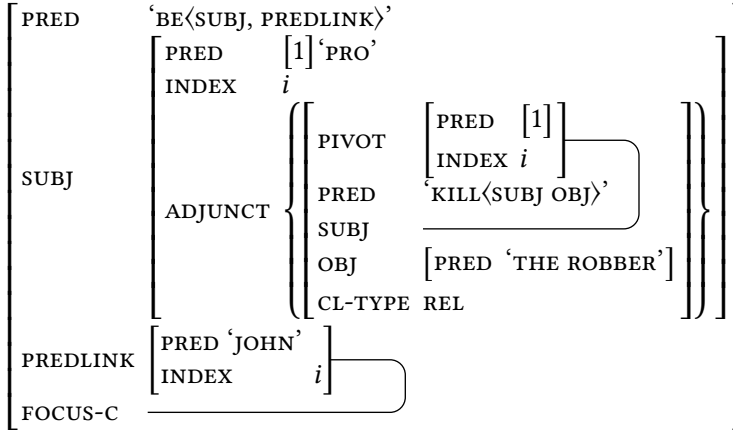
In this analysis, John in (72) is part of the (fronted) PREDLINK (Butt et al. 1999), which is introduced by the copular verb *adalah*.[34] In the absence of *adalah*, the analysis specifies the existence of an unpronounced copular verb in the c-structure. The FOCUS-C in the c-structure is occupied by 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).

---

[34]In 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.

I Wayan Arka & Li-Chen Yeh

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

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-Polynesian
CEMP	Central-Eastern 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 Marking	RC	Relative Clause
DV	Dative Voice	REAL	Realis
GEN	Genitive	REDUP	Reduplication
GR	Grammatical Relation	RELO	Object relativizer
IN	Inchoative	RELS	Subject relativizer
IV	Instrumental Voice	SVC	Serial Verb Construction
LNK	Linker	UV	Undergoer Voice
LV	Locative Voice	WMP	Western 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.

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.

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. Forthcoming. Clause structure and configurationality. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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.

Arka, I Wayan. 2003b. Voice systems in the Austronesian languages of Nusantara: Typology, symmetricality 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.

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.

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.

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.

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.

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.

Arka, I Wayan & Li-Chen Yeh. Forthcoming. LFG and Austronesian languages. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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. Forthcoming(a). Grammatical functions: Their classification and properties. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Belyaev, Oleg. Forthcoming(b). Introduction to LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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.

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. Maxwell III. 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.

Bowden, John. 2001. *Taba: Description of a south Halmahera language*. Canberra: Pacific Linguistics.

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).

Bugenhagen, Robert D. 1995. *A grammar of Mangap-Mbula: An Austronesian language of Papua New Guinea.* Canberra: Australian National University. (Doctoral dissertation).

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.

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.

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. Maxwell III. 2003. Complex predication via restriction. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '03 conference*, 92–104. Stanford: CSLI Publications.

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.

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.

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.

De Guzman, Videa 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.

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.

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. & Anna Kibort. Forthcoming. Argument structure and mapping theory. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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 L. & 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 (ed.). in preparation. *Enggano: Historical and contemporary perspectives.*

Heusinger, Klaus Von. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19. 245–274. DOI: 10.1093/jos/19.3.245.

Hill, Deborah. 2002. Longgu. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic languages*, 538–561. Richmond, Surrey: Curzon Press.

Himmelmann, Nikolaus P. 2005. The Austronesian language 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 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).

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.

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).

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.

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.

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.

King, Tracy Holloway. 2005. Cliticizing LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '05 conference.* Stanford: CSLI Publications.

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 Zainil & Syofyan Adam. 1978. *Bahasa mentawai.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.

I Wayan Arka & Li-Chen Yeh

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).

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*. (Doctoral dissertation).

O'Connor, Rob. 2004. Clitics and phrasal affixation in constructive morphology. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '04 conference*, 315–332. Stanford: CSLI Publications.

Palmer, Bill. 1999. *A grammar of the Kokota language, Santa Isabel, Solomon Islands*. Sydney: University of Sydney. (Doctoral dissertation).

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 society (Volume 1: Material Culture)*, 173–209. Canberra: Pacific Linguistics. DOI: 10.15144/PL-C152.173.

Rákosi, György. Forthcoming. Anaphora. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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.

Sadler, Louisa. Forthcoming. LFG and Semitic languages. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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 Wooi: An Austronesian language of Yapen Island, Western New Guinea*. Canberra: Australian National University. (Doctoral dissertation).

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. Forthcoming. LFG and Minimalism. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

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.

Teng, Stacy Fang-Ching. 2008. *A reference grammar of Puyuma*. Canberra: Pacific Linguistics.

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. Forthcoming. Raising and control. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Wei, Ting-Chi. 2009. 阿美語疑問詞研究 [Wh-interrogatives in Amis]. *Language and Linguistics* 10(2). 315–374.

I Wayan Arka & Li-Chen Yeh

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.

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 symmetricality and syntactic ergativity in Pazeh. In Miriam Butt, Tracy Holloway King & Ida Toivonen (eds.), *Proceedings of the LFG '19 conference*, 415–430. Stanford: CSLI Publications.

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. Forthcoming. Information structure. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Chapter 4

LFG and Celtic languages

Jenny Graver

University of Oslo

This chapter presents an overview of LFG studies on grammatical phenomena in two of the Celtic languages, Irish and Welsh. While there is less work on the Celtic languages in LFG compared to other theories, the studies we have touch on important topics in any linguistic theory or language study, such as word order, grammatical functions, agreement and verbs of existence. The chapter is structured accordingly, and discusses issues such as the presence or absence of a VP, impersonal and passive verb forms, relative clauses and unbounded dependencies, verbal agreement, and the syntax of the Irish copula verb. The Celtic languages are minority languages, and the chapter is framed by reflections on the challenges inherent in studying languages in that situation.

1 Introduction

1.1 The Celtic languages

Historically the Celtic languages are divided into Continental Celtic and Insular Celtic. For the Continental Celtic languages such as Gaulish and Celtiberian, very little is attested. Insular Celtic is normally divided into two branches, the Gaelic or Goidelic group containing Irish, Scottish-Gaelic and Manx, and the British or Brythonic group consisting of Welsh, Breton and Cornish. The Goidelic and Brythonic languages are sometimes referred to as Q Celtic and P Celtic respectively, reflecting the development of Indo-European */k^w/ into /k/ in the Goidelic languages and /p/ in the Brythonic languages (Schmidt 2002: 68).

All the modern-day Celtic languages are minority languages influenced by the strong presence of either English or French as the majority language. While there

are movements to revive Cornish and Manx, these languages have no known traditional native speakers alive.

In minority languages like these, with potentially more speakers who are second language learners than there are native speakers, it is important for a researcher in any linguistic field to be aware of which variety of the language she is working with. For example, Irish is estimated to have 141,000 L1 users and 1,030,000 L2 users (Eberhard et al. 2019). This means that there are for all intents and purposes two Irish language communities, the rural communities of the official Irish-speaking areas called the *Gaeltacht* (plural *Gaeltachtaí*), and urban communities of second-language learners who go on to raise their children in what seems to be developing into new varieties of the language. McCloskey (2003) describes some of the issues involved in working with Irish in this situation. As McCloskey points out, even the question of which variety to study for the purpose of theoretical syntax is fraught with the potential to be felt painfully by the speakers in question. Kennard's (non-LFG) studies on Breton word order (Kennard 2014), and an impersonal construction and initial mutation[1] in Breton (Kennard 2019), are other, excellent examples of some of the complexities involved in studying languages in this situation.

Another issue to be aware of is the differences between the spoken and literary varieties in these languages, a distinction which is particularly prominent in Welsh, but also relevant for Irish. Areas where different varieties come into play in this chapter are among others Irish verbal agreement (dialect, register and diachronic development; Section 4.2) and Irish numerals ("school" language vs. spoken language, Section 4.3).

1.2 On the selection of topics in this chapter

Relatively little work has been done on the Celtic languages in LFG compared to in other theories, and the studies we have cover very different topics. It has been my goal to write an overview chapter that shows some of this breadth. This means that there has not been sufficient room to present all the relevant theory or all the relevant language structures in detail. References to theoretical and grammatical resources are provided, including to other chapters in this Handbook. I encourage the reader to consult the referenced works.

Often the works presented in this chapter are single studies on a single grammatical phenomenon in a single language. What do these studies contribute to our understanding of LFG and of the Celtic languages? What is the theoretical context of the study? I highlight where there remains work to be done in

---

[1]See Section 1.2.

LFG through comparisons with studies on the Celtic languages in other frameworks, and through introducing relevant grammatical phenomena in the Celtic languages that are still unaccounted for in LFG. It is my hope that this may be useful for researchers down the road who want to help fill the gaps in our LFG-theoretical understanding of the Celtic language family, or otherwise study the Celtic languages within the framework of LFG.

The system of initial mutation is a striking example of a central phenomenon in the Celtic languages that has received little attention in LFG. All the Celtic languages have a system of initial mutation in which phonological changes to the initial segment of words are triggered by lexical, morphosyntactic or syntactic conditions. Taking Irish as an example, there are two initial mutations in the language, called lenition and eclipsis. Some examples of how these mutations affect consonants are provided below. In these examples the acute accent denotes a palatalised as opposed to a velarised consonant, called "slender" and "broad" respectively in traditional grammars.

(1) Some initial mutations in Irish, spelling and pronunciation (Mac Eoin 2002: 109)

	Radical		Lenited		Eclipsed
c	/k/, /k′/	ch	/x/, /ç/	gc	/g/, /g′/
d	/d/, /d′/	dh	/ɣ/, /j/	nd	/N/, /N′/
f	/f/, /f′/	fh	(silent)	bhf	/w/, /v′/
s	/s/, /s′/	sh	/h/, /ç/	N/A	

Initial mutation is perhaps one of the most studied Celtic phenomena in general (see Harlow 1989, Ball & Müller 1992, Tallerman 2006 among many others). This might be one reason why it is hard to find LFG studies on this topic beyond computational approaches such as Mittendorf & Sadler's (2006) analysis of Welsh initial mutation using the XLE grammar development environment and the associated finite state and tokenisation tools. However, initial mutation is frequently mentioned when it interacts with the grammatical phenomenon under discussion, such as the Irish relative sentences discussed in Section 3.4.

2 Word order

2.1 Introduction

The Celtic languages show basic VSO word order. As pointed out for example by Fife (2002: 16), the Celtic languages are VSO not only in terms of basic word order

Jenny Graver

– they also show the features proposed by Joseph Greenberg to be implications of basic VSO word order: they are prepositional; they can be said to show SVO as an alternate order through fronting of non-verbal constituents, possibly more correctly described as XVO; they have initial interrogative particles, pre-verbal *wh*-words and the main verb after the auxiliary; and as a main rule, they show post-head modifications.

This section is centered on an LFG analysis of basic VSO word order at the clausal level, with the main issue being the presence of a VP in the Celtic languages. As will be seen, Sadler (1997) and Bresnan (2001) analyse the VSO word order of Welsh in order to develop and illustrate some very central concepts of LFG.

2.2 Is there a VP or not?

A central theoretical discussion concerning the Celtic languages has been the presence or absence of a VP. Early work on this question in other theories than LFG include Sproat (1985) for Welsh, McCloskey (1983) for Irish and Anderson & Chung (1977) for Breton.

As previously mentioned, the Celtic languages show various surface word orders in addition to VSO in different types of clauses. Tallerman (1998: 22–23) distinguishes between what she calls "two major word order patterns in finite clauses in Celtic". The first pattern has the finite lexical verb in initial position followed by the subject, object and any optional material – in other words, the standard VSOX order, as illustrated in (2) for Welsh and (3) for Irish.

(2) Welsh (Tallerman 1998: 23)
     Rhoddais   i afal   i'r   bachgen ddoe.
     give.PST.1SG I apple to.DEF boy   yesterday
     'I gave an apple to the boy yesterday.'

(3) Irish (Ó Siadhail 1989: 205)
     Labhrann Mícheál Gaeilge le Cáit go minic.
     speak.PRS Mícheál Irish   to Cáit often
     'Mícheal often speaks Irish to Cáit.'

The other unmarked word order referred to by Tallerman (1998: 22–23) is a periphrastic construction with an initial finite auxiliary verb, followed by the subject and a non-finite verb and its complement, followed by any optional material. Examples are provided in (4) for Welsh and (5) for Irish, both of which illustrate the progressive construction.

(4) Welsh (Tallerman 1998: 23)

Mae   o'n   adeiladu tai   ym Mangor.
be.PRS he.PROG build   houses in Bangor
'He's building houses in Bangor.'

(5) Irish (Mac Eoin 2002: 131)

Tá   mé ag   baint fhéir.
be.PRS I PROG cut   grass.GEN
'I am cutting grass.'

It is possible to front the non-finite verb and its complement in this construction using the cleft construction, illustrated in (6) for Welsh.

As we will see, this is taken as one indication of the presence of a VP in Welsh.

(6) Welsh (Bresnan 2001: 128)

Adeiladu tai   ym Mangor a   wnaeth   o.
build   houses in Bangor REL do.PST.3SG he
'He built houses in Bangor.' (VP focus)

(6) shows the periphrastic construction with the finite auxiliary verb 'do'; the non-finite verb 'build' and its complement is fronted. Similar fronting is found in Irish, as shown with the periphrastic construction in (7):

(7) Irish (from McCloskey 1983, quoted in Carnie 2005: 14):

Má's ag   cuartughadh leanbh do   dhearbhrathra a   tá   tú ...
if.COP PROG seek   child your brother   REL be.PRS you ...
'If it's seeking your brother's child that you are ...'

Another argument frequently posited in favour of a VP in VSO languages is the presence of structure-dependent subject/object asymmetries such as anaphoric binding (Carnie & Guilfoyle 2000: 5–6 and references therein). The examples in (8) illustrate anaphoric asymmetries for Welsh:

(8) Welsh (Borsley 2006: 476)

a. Welodd   Gwyn ei   hun.
    see.PRS.3SG Gwyn 3SG.M self
    'Gwyn saw himself.'

b. *Welodd   ei   hun Megan.
    see.PRS.3SG 3SG.F[2] self Megan
   Intended: 'Megan saw herself.'

Jenny Graver

If, as in LFG, binding constraints are taken to be a matter for f-structure, examples such as the above are, however, not an argument in favour of a VP in VSO languages.

Based on examples such as (6) and (7) above, Bresnan (2001: 126–131), in line with Sadler (1997), argues in favour of a VP for Welsh as shown in the trees in (9).[3] For Bresnan (2001: 126ff), this argument is a matter of showing an example of what she calls "the noncompositionality of f-structures in c-structures", or more specifically for Welsh and the other Celtic languages that a finite VP can be discontinuous and with a head appearing external to the rest of the phrase. Crucially, this places the analysis of the word order of the Celtic languages in the context of central LFG concepts and analyses such as structure-function mapping and endocentricity and extended heads (see Belyaev forthcoming [this volume]).
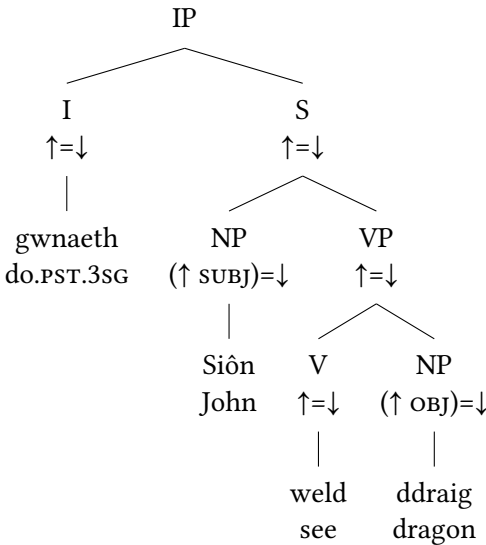
(9)   Welsh word order

   a.   'John saw a dragon.' (Broadwell 2005: 2)
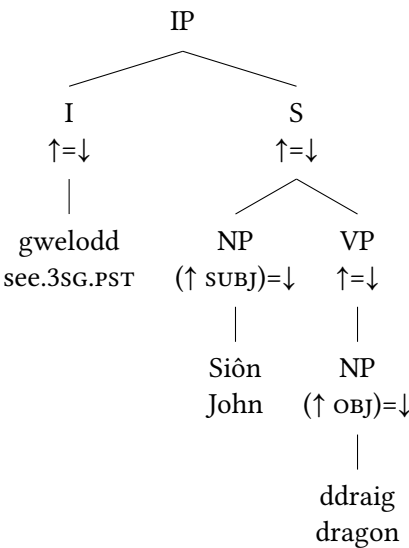
   b.   'John saw a dragon.' (Bresnan 2001: 128)

---

[2]This is glossed 'm' in Borsley (2006: 476).

[3]See Carnie (2005) for a discussion of Irish copula clauses as a possible counter-argument to this type of analysis of VSO languages.

```
                              IP
                    ┌──────────┴──────────┐
                    I                     S
                   ↑=↓                    ↑=↓
                    │              ┌───────┴───────┐
                 gwnaeth          NP              VP
                 do.PST.3SG     (↑ SUBJ)=↓        ↑=↓
                                   │        ┌──────┴──────┐
                                 Siôn       V            NP
                                 John      ↑=↓       (↑ OBJ)=↓
                                           │             │
                                         weld         ddraig
                                          see          dragon
```

In this analysis the V in the I(nfl) position is the extended head of the VP. The tree in (9)a illustrates standard VSO order, whereas (9)b shows a Welsh periphrastic construction with the finite auxiliary verb 'do' in the initial position followed by the subject, a non-finite verb and its complement.

More broadly this analysis deals with several very central questions in any linguistic theory: what is the status of the VP? Is it desirable to maintain a unified analysis of different constructions in a language? Compare for example Borsley (2006), who argues against a head-raising account/discontinuous VP for finite, non-periphrastic clauses in Welsh. Borsley acknowledges the possibility of a VP in periphrastic constructions, but argues that it does not follow that there is a VP in finite, non-periphrastic clauses.

There is much more work to be done on the word order of the Celtic languages, both in general and in LFG, for example in light of Breton apparently showing verb-second effects (Schafer 1995, Tallerman 1998: 22, Stephens 2002: 400) and the development of SVO vs. V2 structure in modern Breton (Timm 1989, Kennard 2014, etc). See also Sadler (2006: 1779–1783) for a discussion of Welsh constituent structure.

2.3 Some other patterns of word order

Within LFG there are so far relatively few studies of Celtic word order beyond clausal VSO structure. One important exception is Sadler's (1998) article "Welsh

Jenny Graver

NPs without head movement", on the structure of Welsh noun phrases. Her starting point is the similarity between nominal and clausal structure in Welsh, which has led to head movement type of analyses of Celtic noun phrases in which the head N raises to a functional category position, parallel to the extended head analyses described for VSO clauses. Sadler argues against this type of analysis for Welsh noun phrases for both conceptual and empirical reasons. She proposes instead an analysis in which Welsh nouns lack complements, and that what appears to be complements of the noun, are instead adjuncts. This removes the need for a head raising account for this data.

These similarities between NPs and VPs in the Celtic languages are another area that would benefit from further study. The issues raised are broader than the Celtic languages; as Sadler (1998: 2) points out, the head raising account of Celtic noun phrases is modelled on analyses of Semitic noun phrases, which show similarities with the Celtic structures.

In the introduction to this section, I mentioned some salient typological features of the Celtic languages that correspond to a VSO word order. One of these is fronting, or clefting. The Celtic pattern of fronting (see Tallerman 1998: 31–34 etc) is illustrated below using Irish and Welsh. The basic structure of the Irish cleft construction is copula + clefted phrase + relative particle + the remainder of the sentence. This is illustrated in (10) through (a) a standard VSO sentence, (b) fronting of the subject, and (c) fronting of an adverb:

(10) Irish (Sulger 2009: 571):

    a. Léigh   an múinteoir leabhar inné.
       read.PST DEF teacher   book   yesterday
       'The teacher read a book yesterday.'

    b. Is é   an múinteoir a   léigh   an leabhar inné.
       COP AGR DEF teacher   REL read.PST DEF book   yesterday
       'It is the teacher who read a book yesterday.'

    c. Is inné   a   léigh   an múinteoir an leabhar.
       COP yesterday REL read.PST DEF teacher   DEF book
       'It is yesterday that the teacher read a book.'

In Welsh, the corresponding construction does not have a copula, leaving the relative particle as the only marker of clefting (Watkins 2002: 336–337):

(11) Welsh (Watkins 2002: 337)
    y   bachgen a   welodd   y   dyn
    DEF boy   REL see.PST.3SG DEF man
    'It was the boy who saw the man.'

The cleft construction in the Celtic languages has received relatively little attention in LFG, but see Sulger (2009) for an analysis that builds on analyses of the Irish copula.[4]

3 Arguments

3.1 Introduction

This section starts with an analysis of the Modern Irish so-called "autonomous" verb form and its diachronic development from a passive verb, which is then contrasted briefly with the Welsh impersonal verb form. These analyses deal with a crucial topic in any grammatical theory, namely mapping between verbal semantics and syntactic functions. The autonomous verb is followed by a description of a pattern in Welsh in which an adjective phrase is said to select for an object. The authors in question argue that this analysis raises wider issues about how best to understand grammatical functions in areas outside of verbal subcategorisation. Finally, there is a brief discussion of Irish relative clauses in the context of LFG analyses of unbounded dependencies.

3.2 Passives and impersonals

All the Celtic languages contain a verb form in their paradigm called "autonomous" or "impersonal" (Fife 2002: 14). There are two PhD theses dealing with this verb form within the framework of LFG, Graver (2010) for the Irish autonomous verb and Arman (2015) for the Welsh "impersonal passive" as well as another type of Welsh passive called the GET-passive. Both make use of Lexical Mapping Theory (LMT) as revised by Kibort (2007, 2014) (see Findlay & Kibort forthcoming [this volume]).

Some classic studies of the Modern Irish autonomous verb are Stenson (1989) and McCloskey (2007). Their main conclusion is that the Modern Irish autonomous verb is an active verb with an impersonal subject comparable in semantics to French *on* etc. In Irish, this subject is phonologically null. Drawing on this conclusion, Graver (2010, 2011) presents an LFG analysis of the Modern Irish autonomous verb and its diachronic development.

---

[4]Compare also Borsley (2020), a comparative analysis in HPSG of *wh*-interrogatives, free relatives and cleft sentences. Borsley suggests that the Welsh cleft construction involves identity predication. As mentioned, the copula does not appear in Welsh cleft sentences today. This leads Borsley to suggest that the identity predication is associated with the construction in Modern Welsh, whereas in Middle Welsh, where the copula did appear, the identity predication of the cleft sentence was associated with the copula.

Jenny Graver

Examples of the Modern Irish autonomous verb are provided in (12):

(12) Irish
  a. Tugadh   an corp chun na reilige   agus cuireadh   é.
      bring.PST.AUT DEF corpse to   DEF graveyard and   put.PST.AUT it
      'The corpse was brought to the graveyard and it was buried.' (Graver
      2010: 4)

  b. Deir   siad go gcuirfear   ar ath-chúirt é.
      say.PRS they that put.FUT.AUT on re-court   it
      'They say that it will be appealed.' (Graver 2010: 9)

As will be shown in Section 4.2, the Modern Irish verbal paradigm contains a mix-
ture of so-called synthetic forms, which express person and number, and analytic
forms, which are used with separate pronouns. The autonomous form can thus
be interpreted as a synthetic form expressing a subject with impersonal meaning,
similar to a third person singular subject, etc.

The agent phrase is ungrammatical with the autonomous verb. This is an ar-
gument in favour of an active, impersonal analysis instead of a passive analysis.
Assuming an analysis of the agent phrase as an oblique rather than an adjunct,
this ungrammaticality is predicted by analysing the autonomous verb as an ac-
tive, synthetic form, since the first argument of the verb is mapped to the im-
personal subject and is thus unavailable for mapping to the agent phrase.[5] The
ungrammaticality of the agent phrase with an autonomous verb is illustrated in
(13).

(13) Irish (Stenson 1989: 382)
      *buaileadh   Ciarraí {ag, le}   Gaillimh
      beat.PST.AUT Kerry   by, with Galway
      Intended: 'Kerry was beaten by Galway.' [in a hurling match or similar]

Another argument in support of the same conclusion is object marking on the
patient argument (*é* 'it' in (13)). Stenson illustrates this as follows, showing the
ungrammaticality of the subject pronoun *siad* 'they' instead of the object pro-
noun *iad*:

(14) Irish (Stenson (1989: 384)

---

[5]See Graver (2010: 60–61 and references therein) for arguments in favour of analysing the agent
  phrase as an oblique as opposed to an adjunct.

234

buaileadh   aríst iad/*siad
beat.PST.AUT again them/*they
'They were beaten again.'

What is more, the autonomous verb may be used with more or less all verbs including intransitive and unaccusative verbs. I refer to the abovementioned references for additional data in favour of an active impersonal analysis.

Kibort (2007) reformulates the principles for mapping between arguments and grammatical functions compared to classic LMT, and suggests the following mapping principle.

(15) Mapping principle (Kibort (2007): 16)
      The ordered arguments are mapped onto the highest (i.e. *least* marked) compatible function on the markedness hierarchy. [emphasis original]

The markedness hierarchy referred to in (15) is the classic hierarchy provided in (16), which again is based on the feature decomposition of f-structure functions shown in (17):

(16) Partial ordering of syntactic functions in terms of markedness (Bresnan (2001): 309)

$$\text{SUBJ} \; < \; \text{OBJ, OBL}_\theta \; < \; \text{OBJ}_\theta$$

(17) Feature decomposition of f-structure functions (Bresnan 2001: 308)

	$[-r]$	$[+r]$
$[-o]$	SUBJ	OBL$_\theta$
$[+o]$	OBJ	OBJ$_\theta$

Thus, the mapping between a- and f-structure will simply look as follows for a transitive autonomous verb, where *impers* is shorthand for "impersonal" – this mapping is similar to a regular active, transitive verb with any kind of subject.

(18) Mapping, transitive autonomous verb (Graver 2010: 62)

verb [aut.trans.] ⟨   arg1   arg2   ⟩
                          $[-o]$ $[-r]$
                          $[-r]$ $[+o]$
                    ─────────────────────────
                     SUBJ*impers*   OBJ

Three different morphosyntactic operations can be formulated to account for passivisation with and without an agent phrase, and with the second argument

mapped to either the subject function (canonical passive) or the object function (impersonal passive). All of these operations result in passive verbs, the mapping of which is incompatible with the analysis of the autonomous verb as active with an impersonal subject as shown in (18).

When the agent phrase is not present in a passive sentence, arg1 undergoes mapping to zero/∅ (Bresnan 2001: 310). When the agent phrase is present, it undergoes mapping to oblθ (Kibort 2007: 17–19). In a canonical passive, arg2 maps to subj as the least marked compatible function. In an impersonal passive, an operation called object preservation applies to map arg2 to obj, which entails an increase in the markedness of the mapping.

Where these mapping relations really turn out to be of use according to Graver (2010, 2011) is in the analysis of the diachronic development of the Modern Irish autonomous verb. In Old Irish, the properties of the autonomous verb appear contradictory in terms of the above mappings. Graver (2010: 179) sums this up as follows:

(19)   Properties of the Old Irish autonomous verb:

   a.   It is found included in the paradigm for practically any verb in every category of tense/aspect/mood, including intransitive and unaccusative verbs such as the substantive verb (Section 5.1) and verbs of inherently directed motion. (See Graver (2010: 62–63 and references therein) on passivisation and unaccusativity in LFG in the context of the Irish autonomous verb.)

   b.   A third person patient is marked as subject, by nominative case on nouns, agreement in number with the verb and by the verb itself if the patient is a pronoun.

   c.   There is object marking on first and second person patients, with infixed pronouns.

   d.   The agent phrase is possible with transitive verbs.

The development from the above situation to the Modern Irish active impersonal can be summarised in terms of the markedness inherent in Kibort's (2007) theory: due to general changes in the morphological system of the Irish language, the patient of the Old Irish "passive" verb is reanalysed as the object rather than the subject of the verb. The resulting impersonal passive is predicted by the theory to be more marked than the original canonical passive, since an additional morphosyntactic operation/increase in markedness, object preservation, has applied.

The resulting subjectless, impersonal passive can be considered an unstable category, and for example Blevins (2003: 480–481 and references therein) suggests that subjectless impersonal passives tend to have an indefinite human agent interpretation, and that a subjectless impersonal passive thus would be practically indistinguishable from an active impersonal and consequently susceptible to reanalysis. When the autonomous verb in Irish is reanalysed as containing an impersonal active subject, there is no longer any need for the morphosyntactic increase in markedness – but see Graver (2010: 200–203) for a discussion of the difficulties of pinpointing the exact causes of such a diachronic change.

In other words, the status of the Old Irish autonomous verb could be termed contradictory or unclear in terms of the morphosyntactic operations illustrated above and the resulting impersonal and passive constructions. A comparable situation appears to apply in Modern Welsh: for example, the Welsh impersonal verb form can occur both with an agent phrase and unaccusative verbs. This phenomenon is analysed in terms of LFG by Arman (2015). Arman does not conclude whether the Welsh "impersonal" verb is in fact passive or active, but suggests that LFG, and particularly the revised mapping theory, is flexible enough to account for the Welsh data (Arman 2015: paragraph 7.3). The strength of Arman's approach is the large amount of data, the comparisons with other passives in the language and, in particular, the detailed analysis of the interaction of the impersonal verb form with different semantic verb classes (chapter 6). A similar LFG analysis of the autonomous verb in Old Irish would be highly interesting.

3.3 A Welsh adjectival construction

Mittendorf & Sadler (2008) analyse a Welsh adjective phrase construction containing a noun phrase as a constituent. They call this the *in-respect-of* construction, illustrated in (20), where the adjectives *byr* 'short' and *trwm* 'heavy', respectively, are followed by noun phrases containing a possessive clitic pronoun, here *ei* 'her':

(20) Welsh (Mittendorf & Sadler 2008: 2)

    a. byr   ei   thymer
        short her temper
        'short-tempered'

    b. trwm ei   chlyw
        heavy her hearing
        'hard of hearing'

Jenny Graver

Mittendorf & Sadler (2008: 19–20) suggest that the main theoretical contribution of their analysis of this Welsh construction is a call for more specific descriptions of the grammatical functions of LFG, particularly outside of the area of verbal subcategorisation.

Mittendorf & Sadler show that the construction occurs in similar environments to adjective phrases, as shown in (21) for attributive and predicative use respectively.

(21) Welsh (Mittendorf & Sadler 2008: 9)

    a. merch fyr   ei thymer
       girl   short her temper
       'a short-tempered girl'

    b. Mae'r   ferch yn fyr ei thymer.
       be.PRS.DEF girl   PRED short her temper
       'The girl is short-tempered.'

They go on to provide evidence, following apparently unpublished work by Jones (2002), in favour of analysing the adjective-NP sequence as one constituent, which is headed by the adjective. Phenomena in favour of this analysis include coordination – the NP in the sequence can be coordinated, which indicates that it is a subconstituent (Mittendorf & Sadler 2008: 3) - and the way that the adjective in the sequence can be modified as expected by regular adverbials and other types of intensifiers.

Initial mutation occurs in Welsh on an adjective modifying a feminine singular noun. The *in-respect-of* construction behaves as expected for an adjective when it modifies a singular feminine noun, as illustrated in (22), where the adjective *mawr* 'big' becomes *fawr* following the feminine singular noun *athrawes* '(female) teacher':

(22) Welsh (Mittendorf & Sadler 2008: 6)
    athrawes fawr ei   pharch
    teacher big   her respect
    'a highly-respected (female) teacher'

Having established that the adjective-noun sequence is a constituent headed by the adjective, Mittendorf & Sadler's main question is: What is the correct f-structure analysis of the noun phrase contained in the adjective phrase? What is its grammatical function? They review and reject analyses in which the noun

phrase is a SUBJ and ADJUNCT, and tentatively conclude that the noun phrase is an OBJ.

They provide the f-structures in (23) for attributive and predicative use of the construction (Mittendorf & Sadler 2008: 12):

(23) F-structures, the *in-respect-of* construction (Mittendorf & Sadler 2008: 12)

a. attributive use:

$$
\begin{bmatrix}
\text{PRED} & \text{`GIRL'}_i \\
\text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{`SHORT}\langle\text{OBJ}\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`TEMPER}\langle\text{POSS}\rangle\text{'} \\ \text{POSS} & [\text{PRED } \text{`PRO'}_i] \end{bmatrix} \end{bmatrix} \right\}
\end{bmatrix}
$$

b. predicative use:

$$
\begin{bmatrix}
\text{PRED} & \text{`SHORT}\langle\text{SUBJ, OBJ}\rangle\text{'} \\
\text{SUBJ} & [\text{PRED `GIRL'}] \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`TEMPER}\langle\text{POSS}\rangle\text{'} \\ \text{POSS} & [\text{PRED } \text{PRO}_i] \end{bmatrix}
\end{bmatrix}
$$

Mittendorf & Sadler suggest that it might sound surprising that an adjective selects for an object, but propose that this is a reasonable analysis given the resources of the theoretic arsenal of LFG, as well as some cross-linguistic support from Swedish among other languages (Mittendorf & Sadler 2008: 18). However, their main argument in support of an OBJ analysis of the noun in this construction is a comparison with a very similar Welsh *tough* construction and the mandatory presence of the noun. See Kaplan & Zaenen forthcoming: §5 [this volume] on the *tough* construction in English in the context of long-distance dependencies. The *tough* construction in Welsh is illustrated below, with a verbal noun as the COMP of the adjective *treulio*, verbal noun of 'to digest' in the example below, and the mandatory presence of the noun, which argues against an ADJUNCT analysis.

(24) Welsh *tough*-construction (Mittendorf & Sadler 2008: 14)
bwyd anodd   ei dreulio
food difficult its digest
'food difficult to digest'

3.4 Unbounded dependencies – the Irish relative clause

What happens when an argument of the verb is taken out of its normal position through relativisation? Irish has two relativisation strategies that conform with

Jenny Graver

the Accessibility Hierarchy (Keenan & Comrie 1977). One of the earliest descriptions of these facts is McCloskey (1979). For an analysis of the more complicated Welsh data, see for example Tallerman (1990) and Borsley (2013).

In traditional grammar (such as The Christian Brothers 2002), the two Irish relativisation strategies are called the direct and the indirect relative. The direct relative is a gap strategy, whereas the indirect relative uses a resumptive pronoun.

The direct relative is used when the relative constituent is the subject or the object. It uses a relative particle that lenites.[6]

(25) Irish: the direct relative (McCloskey 1979: 5–6)

    a. Relativised subject
       an fear a   dhíol an domhan
       DEF man REL^L sell.PST DEF world
       'the man who sold the world'

    b. Relativised object
       an scríbhneoir a   mholann na mic léinn
       DEF writer   REL^L praise.PRS DEF students
       'the writer whom the students praise'

The direct relative is obligatory with a relativised subject and the most common with a relativised object (McCloskey 1979: 6). However, since the VSO word order gives rise to potential ambiguity in examples like (25)b, the indirect relative with a resumptive pronoun is possible in these cases:

(26) Irish: indirect relative with a relativised object (McCloskey 1979: 6)
      an scríbhneoir a   molann na mic léinn é
      DEF writer   REL^N praise.PRS DEF students him
      'the writer whom the students praise'

Going further down the Accessibility Hierarchy, the Indirect Relative is obligatory with objects of prepositions and possessors:

(27) Irish: the indirect relative (McCloskey 1979: 6)

---

[6]Lenition and eclipsis are the two initial mutations in Irish, as explained in Section 1.2. They are glossed L for lenition and N for nasalisation, the latter a traditional – but imprecise – term for eclipsis.

a. Relativised prepositional object
an fear a dtabharann tú an t-airgead dó
DEF man REL[N] give.PRS you DEF money to.him
'the man to whom you give the money'

b. Relativised possessor
an fear a bhfuil a mháthair san otharlann
DEF man REL[N] be.PRS his mother in.DEF hospital
'the man whose mother is in the hospital'

These core Irish facts, as well as some more peripheral patterns described by McCloskey (2002), are analysed in detail by Asudeh in his book *The Logic of Pronominal Resumption* (Asudeh 2012: chapter 7). The chapter on long distance dependencies in this volume (Kaplan & Zaenen forthcoming [this volume]) provides a brief description of Asudeh's analysis in the context of the development of LFG analyses of the explicit marking of f-structures in the domain of a long distance dependency – though it should be noted that Kaplan restricts himself to examples with the direct relative, in comparison with sentences with the Irish complementiser *go[N]* 'that' – which introduces complements that are not in the domain of a long distance dependency.

4 Agreement

4.1 Introduction

There are many agreement issues in the Celtic languages that remain untouched within the LFG framework. In this section I present analyses of Irish verbal conjugation and a Welsh conjunct agreement pattern, before moving on to agreement between cardinal numbers and numerals in Welsh and Irish respectively. There are clear similarities in these areas between the two languages discussed, but also interesting differences that lack thorough analyses.

4.2 Various issues of verbal agreement

Table 1 shows parts of the standard conjugation of the Irish verb *mol* 'to praise'.[7] There is variation between analytic forms, which take a separate pronoun (or noun) subject, and synthetic forms (marked in bold) which are conjugated for person and number.

---

[7]Welsh is quite different from Irish in this respect (see Borsley et al. 2007: 9–10). Welsh shows complete paradigms of synthetic verbal morphology. Literary Welsh permits null subjects; col-

Jenny Graver

Table 1: Irish conjugations (The Christian Brothers 2002: 95) emphasis added)

	Present tense	Past tense	Imperfect
1sg	**molaim**	mhol mé	**mholainn**
2sg	molann tú	mhol tú	**mholtá**
3sg	molann sé/sí	mhol sé/sí	mholadh sé/sí
1pl	**molaimid**	**mholamar**	**mholaimis**
2pl	molann sibh	mhol sibh	mholadh sibh
3pl	molann siad	mhol siad	**mholaidís**

There has been a general development in Irish towards more analytic forms, but as e.g. Ó Siadhail (1989: 182–185) points out, there is a mixture of synthetic and analytic forms in all the dialects with a tendency for the most synthetic forms in the south and the fewest in the north of the country.

There are two important descriptive generalisations associated with Irish verbal agreement. First, as a general rule, the synthetic forms are incompatible with a pronoun or noun subject, as shown in (28).

(28) *molaim   mé
      praise.prs.1sg I

Second, when the paradigm contains a synthetic form, the analytic form is unavailable (though see below for a potential exception).

Irish data such as these are used by Andrews (1990) as a basis for formulating the Morphological Blocking Principle. The main intuition behind this principle is that if there is a highly specified form in the Lexicon, a less highly specified one cannot be used (Andrews 1990: 508).

Andrews (1990) shows first of all that it follows from general LFG architecture that a synthetic verb form cannot occur together with a noun phrase or pronoun.

Synthetic verb forms are taken to specify the value of the PRED of the subject as 'PRO'. A subject NP would contribute a different PRED value than that to the subject, and this is ruled out by the Uniqueness Condition (see Belyaev forthcoming: §3.4.1 [this volume]).

However, as Andrews (1990: 516) points out, the Uniqueness Condition is not sufficient to rule out the presence of a pronoun with a synthetic verb form, since the specification of the PRED value of the subject as 'PRO' from both the verb

---

loquial Welsh does not.

form and the pronoun would not appear to be contradictory. To solve this, Andrews refers to the principle of Predicate Indexing, which he suggests "causes each PRED-value introduced in a lexical item to receive a unique index, which distinguishes it from all other PRED-values in the structure" (Andrews 1990: 516). This principle makes a synthetic verb form and a pronoun subject mutually exclusive, since the 'PRO' values contributed by a synthetic verb form and a subject pronoun respectively would carry separate indices, which in its turn would violate the Uniqueness Condition.

Finally, to account for the ungrammaticality of an analytic verb form with a pronominal subject when there is a synthetic form available, Andrews (1990: 519) formulates the Morphological Blocking Principle:

> Suppose the structure S has a preterminal node P occupied by a lexical item l_1, and there is another lexical item l_2 such that the f-structure determined by the lexical entry of l_1 properly subsumes that determined by the lexical entry of l_2, and that of l_2 subsumes the f-structure associated with P in S (the complete structure, after all unifications have been carried out). Then S is blocked.

Building on this, Sulger (2010) offers a computational LFG analysis of Irish verbal agreement facts. As a part of his analysis, Sulger (2010: 169–170) criticises the Morphological Blocking Principle in computational terms, suggesting that this principle has the consequence that the lexicon needs to be checked for a corresponding synthetic form every time an analytic form occurs. If there is a synthetic form, the analytic form is blocked. Sulger (2010: 170) argues, from a computational grammar viewpoint, that this approach is inefficient and that it is questionable whether it is adequate for larger-scale grammars.

McCloskey & Hale (1984: 491–492 and section 6) point out that there is greater variation in the Irish paradigms than described above (see also Ó Siadhail 1989: 182–185), and that in certain cases the same person-number combination can be expressed both by a synthetic and an analytic form. Some of their examples are included in (29):

(29) Irish (McCloskey & Hale 1984: 491)

    a. chuirfidís
       put.COND.3PL

    b. chuirfeadh siad
       put.COND   they

Jenny Graver

This matter seems to involve both dialect and register variation as well as historical developments, and is so far an understudied topic within the framework of LFG. McCloskey & Hale (1984: 531) indicate morphological blocking as the potentially most fruitful line of enquiry going forward.

Sadler (1999) discusses another agreement phenomenon in Welsh, a single conjunct agreement pattern illustrated in (30):

(30) Welsh (Sadler 1999: 2)

    a. Daeth   Siôn a   minnau.
       come.PST.3SG Siôn and 1SG
       'Siôn and I came.'

    b. Daethost   ti a   minnau/Siôn.
       come.PST.2SG 2SG and 1SG/Siôn
       'You and I/Siôn came.'

    c. Roedd   Mair a fi i briodi.
       be.PST.3SG Mair and 1SG to marry
       'Mair and I were to marry.'

    d. Roeddwn i a   Mair i briodi.
       be.PST.1SG 1SG and Mair to marry
       'I and Mair were to marry.'

All these examples have a plural coordinate subject.[8] When the first conjunct is a pronoun, the verb agrees with the pronoun in person and number. When the first conjunct is non-pronominal, the verb is in the unmarked third singular form. An identical asymmetrical agreement pattern shows up both in nominal structures containing possessor phrases and with objects of prepositions.

Sadler (1999: 3–4) suggests that a similar agreement pattern is found in Irish, based on data from McCloskey (1986). Some of McCloskey's examples are provided in (31):

(31) Irish (McCloskey 1986: 248)

    a. Bhíos   féin agus Tomás ag caint le   chéile.
       be.PST.1SG self and Tomás PROG talk with each.other
       'Tomás and I were talking to one another.'

---

[8]See Sadler (2006) for a discussion of other coordination patterns in Welsh within the framework of LFG.

b. Bhíos   –sa   agus Pádraig Ó Guithín le pósadh.
be.PST.1SG –CONTR and Pádraig Ó Guithín to marry
'Pádraig Ó Guithín and I were to marry.'

Note the elements *féin* and *–sa* here; these are emphatic/contrastive elements that are mandatory in the above coordination pattern.[9]

Andrews (1990: 522–523) mentions this Irish pattern of agreement. He suggests that it presents significant difficulties for an LFG analysis and chooses to leave them aside in the context of his paper on morphological blocking. This type of pattern is also not restricted to the Celtic languages; it appears to be found in for example Czech, Latin and Palestinian Arabic (Sadler 1999: 4 and references therein).

The interesting difficulty with the Welsh data is, as Sadler (1999: 15) puts it, that "morphosyntactic and semantic agreement come apart under coordination": the only difference between the structure illustrated above and other coordinate structures is the agreement between the first, pronominal conjunct and the verb. On the other hand, data such as predicate agreement seem to indicate that semantic feature resolution appears to operate on coordinate structures in Welsh independent of whether the coordinate structure includes pronouns or not.

Sadler (1999) describes two main features of the classic LFG view of agreement: agreement features such as person, number, gender and case are an f-structure phenomenon, and agreement is a matter of constraints on the same structure rather than matching between features on different structures (see Haug forthcoming [this volume] on agreement). The crucial question then is this: can this view of agreement be reconciled with the single conjunct agreement pattern illustrated above? Sadler (1999) argues that these data show that it is difficult to maintain a simple and homogenous view of what agreement is.

4.3 Noun phrase agreement: numerals

Fife (2002: 21) lists as "a common feature of Celtic nominal syntax" the use of singular forms (and/or special forms) following cardinal numerals. Mittendorf & Sadler (2005) make use of the INDEX/CONCORD distinction (referencing Wechsler & Zlatić's (2000) HPSG analysis and King & Dalrymple (2004) in LFG) to account for the resulting agreement mismatch in Welsh noun phrases.

---

[9]See McCloskey & Hale 1984: 493–496 for a thorough discussion of these and other elements and arguments why they are not pronouns.

Jenny Graver

The INDEX/CONCORD distinction describes two sets of nominal agreement features: CONCORD features relate to agreement between the noun and any determiners or adjectives, whereas INDEX features are related to the semantics of the noun and agreement between the noun phrase and a bound pronoun and often also verb agreement (see Haug forthcoming: §3 [this volume] for further details).

In Welsh, numerals require the singular form of the noun, as shown below in the examples 'five dogs' and 'three cats', where the noun in both cases is in the singular form:

(32) Welsh (Mittendorf & Sadler 2005: 6)

    a. pum ci
       five dog.M.SG
       'five dogs'

    b. tair cath
       three.F cat.F.SG
       'three cats'

What is more, if the noun is modified by an adjective with a distinct plural form, the singular form is used. In (33), the adjective *arall* 'other' is used in the plural form *eraill* in the phrase 'other dogs', but in the singular form *arall* when a numeral is added:

(33) Welsh (Mittendorf & Sadler 2005: 6)

    a. cŵn eraill
       dog.M.PL other.PL
       'other dogs'

    b. pum ci arall
       five dog.M.SG other.SG
       'five other dogs'

Demonstratives on the other hand are always plural when a noun with a plural premodifier is involved. In the below examples, the singular (feminine) form *hon* 'this' is used in the phrase 'this cat', whereas the plural form *hyn* is used in 'these three cats':

(34) Welsh (Mittendorf & Sadler 2005: 6)

    a. y gath hon
       DEF cat.F.SG this.F.SG
       'this cat'

> b. y   tair cath   hyn
>   DEF three cat.F.SG this.PL
>   'these three cats'

At the same time, the noun phrase behaves overall as plural, as shown in (35), where the noun phrase 'the five men' controls a pronominal anaphor:

(35) Welsh (Mittendorf & Sadler 2005: 7)
      Roedd   y   pum dyn   yn   gweld eu hunain yn y   drych.
      be.IPFV.3S DEF five man.M.SG PROG see   3PL self.PL in DEF mirror
      'The five men saw/were seeing themselves in the mirror.'

Mittendorf & Sadler (2005) suggest that these distinctions can be most usefully described as an INDEX/CONCORD mismatch: specifically, the numeral contributes the INDEX NUM feature, which will be plural. This accounts for the example in (35), where the noun phrase 'the five men' controls a plural anaphor. What is more, it accounts for the plural demonstrative *hyn* if Welsh demonstratives show INDEX agreement. On the other hand, the singular noun following the numeral will contribute a singular CONCORD NUM feature. This accounts for the requirement that adjectives modifying the noun be in the singular form since adjectives are taken to show CONCORD agreement.

The f-structure for the noun phrase *tri dyn* 'three men' is shown in (36) to illustrate:

(36) Welsh (Mittendorf & Sadler 2005: 11)
      tri   dyn
      three.M man.M.SG
      'three men'

$$
\begin{bmatrix}
\text{PRED} & \text{'MAN'} \\
\text{INDEX} & [\text{NUM PL}] \\
\text{CONCORD} & [\text{NUM SG}] \\
\text{ADJ} & \{[\text{PRED 'THREE'}]\}
\end{bmatrix}
$$

Irish numerals show agreement patterns of the same type as Welsh, in that numerals, as a main rule with certain exceptions described below, are followed by a noun in the singular. How this system interacts with adjective agreement lacks analysis in LFG for Irish.

Describing what he calls the "traditional" system, Ó Siadhail (1982) shows that the main rule also in Irish is that the unmarked, singular form of the noun is used after cardinal numerals:

Jenny Graver

(37)   Irish

    a. trí   chnoc
        three hill.M.SG
        'three hills' (Ó Siadhail 1982: 99)

    b. dhá chnoc   d(h)éag[10]
        two hill.M.SG ten
        'twelve hills' (Ó Siadhail 1982: 100)

    c. trí   chnoc   fhichead
        three hill.M.SG twenty.GEN
        'twenty-three hills' (Ó Siadhail 1982: 101)

The "traditional" system referred to above is a system based on multiples of twenty:

(38)   Irish

    a. deich lá   fichead
        ten   day twenty.GEN
        'thirty days' (Ó Siadhail 1982: 101)

    b. naoi lá   dhéag is   fiche
        nine day ten   and twenty
        'thirty-nine days' (Ó Siadhail 1982: 101)

    c. lá   is   dá fhichead
        day and two twenty.GEN
        'forty-one days' (Ó Siadhail 1982: 102)

However, as Ó Siadhail (1982: 101) points out, what he calls the "school system" has introduced numerals in a decimal system, such as *tríocha* 'thirty', *ceathracha* 'forty', etc.[11] This latter system is considered standard today, and is illustrated in

---

[10]Whether the form has undergone mutation or not (*déag* vs. *dhéag*) is a matter of dialectal variation (see Ó Siadhail 1989: 100).

[11]The terms used by Ó Siadhail when he makes the distinction "traditional" and "school" system, highlight the need for linguists to be aware of the sociolinguistic nuances of the language under study, in order to be certain of which linguistic system we are describing and analysing at a given time. Mac Eoin (2002: 118–119) suggests that the decimal system is in fact a survivor from the literary language, whereas the vigesimal system has prevailed in the spoken language. He goes on to state that '[t]he promotion of the decimal system in the schools during the last seventy years has not diminished the popularity of the vigesimal system in ordinary speech.' (Mac Eoin 2002: 119)

(39) with examples from the school grammar book New Irish Grammar by the Christian Brothers:

(39) Irish (The Christian Brothers 2002: 76)

    a. trí   chapall is   tríocha
       three horse   and thirty
       'thirty-three horses'

    b. seacht gcapall is   caoga
       seven horse   and fifty
       'fifty-seven horses'

The numerals three to ten are however used with certain nouns in the plural. This holds for both number systems. As illustrated in the examples above, the '-teen' part of the numeral phrase – whether the abovementioned multiples of twenty or the school system numerals – is placed after the modified noun while the numbers 1–10 are placed before the noun. Consequently, the exception to the singular rule is relevant for the number system in general and not just when counting to ten.

Nouns used in the plural with numerals can be divided into different groups, including nouns that express a unit of measure (Ó Siadhail 1982: 102–104) and "words inherent to the counting system" such as *ceann* 'head/one' vs. *trí cinn* 'three' (literally 'three heads') (Ó Siadhail 1989: 167)

There is in other words significant variation in the Irish numeral system, depending on whether you are dealing with the traditional or standard written language or the traditional spoken language with its many dialects. We may perhaps also expect to see that the use of the singular form of nouns following numerals is on the way out in the urban varieties of Irish, on the pattern of English.

5 The copula

5.1 Introduction

All the Celtic languages show or have shown a distinction between two 'be' verbs, usually labeled the substantive verb (Irish: *bí*) and the copula (Irish: *is*) (Fife 2002: 19–20 etc.). In LFG it is mainly the Irish copula that has been studied, and thus Irish will be the focus here.[12] This means that Irish copula predication has not

---

[12]Welsh has one copula verb *bod*, which appears to share properties with both the Irish copula and the Irish substantive verb (see Borsley 2019). A comparative LFG analysis of the Irish and Welsh copula systems would be interesting.

Jenny Graver

been studied in its entirety within in LFG. In theoretical terms, the Irish copula and the Irish substantive verb are both copulas, but I will use the traditional labels.

For the Irish copula, it is customary in traditional grammar to distinguish between two types of copula sentences, classificatory and identificatory (e.g. Ó Siadhail 1989: 224). Some examples are provided below, as context for the following theoretical discussion:

(40) Irish (Ó Siadhail 1989: 224)

    a. Is   scoláire mé.
       COP scholar I
       'I am a scholar.'

    b. Is   múinteoir í   Cáit.
       COP teacher   AGR.3SG.F Cáit
       'Cáit is a teacher.'

(41) Irish (Ó Siadhail 1989: 227)

    a. Is   mé an múinteoir.
       COP I   DEF teacher
       'I am the teacher.'

    b. Is   é   Seán an múinteoir.
       COP AGR.3SG.M Seán DEF teacher
       'Seán is the teacher.'

In classificatory sentences such as those in (40), the subjects *mé* 'I' and *Cáit* are said to belong to the class of scholar/teacher. The identificatory sentences in (41) express identity between the subjects, *mé* 'I' and *Seán*, and 'the teacher'.

In this section I first discuss the syntax of the Irish copula. There are two main types of analysis proposed in the LFG literature for copula constructions, a single-tier analysis where the PRED of the sentence is the non-verbal predicate, and a double-tier analysis with two varieties depending on the choice of argument function for the non-verbal predicate. It is shown that while LFG works on the Irish copula tend towards a double-tier, PREDLINK analysis, there is philological work on older stages of the language that suggest a single-tier analysis as more appropriate to the Irish data. I then go on to show how the Irish copula behaves in terms of the distinction between stage level and individual level.

5.2 Syntax of the Irish copula

In the LFG literature (Dalrymple et al. 2004 and references therein), there are three types of analyses suggested for different types of copula constructions across languages, as shown in Figure 1.
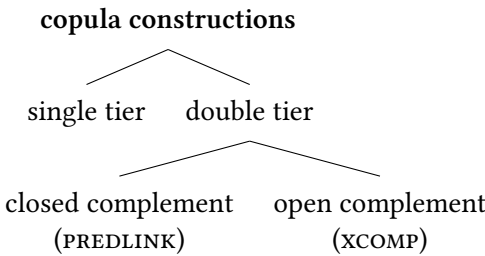
copula constructions

single tier   double tier

closed complement   open complement
(PREDLINK)   (XCOMP)

Figure 1: Types of copula constructions (adapted from Sulger 2009: 564)

Dalrymple et al. (2004) suggest that different f-structure analyses are appropriate for different copula constructions not only between languages but also within in a single language. Attia (2008) on the other hand argues in favour of a unified, general analysis of copula constructions on the f-structure level, and suggests that the variations in morphological agreement, presence or absence of the copula, etc, used as arguments in favour of different analyses by Dalrymple et al., do not warrant functional variation.

Sulger (2009) mostly follows Attia (2008) and argues that a PREDLINK analysis is universally applicable to copula constructions, thus also for Irish. In the following I show how the Irish data have been situated in the context of this discussion. I will briefly sketch the three types of copula analyses as context for Sulger (2009) analysis, before providing his main arguments in favour of a double-tier, PREDLINK analysis for the Irish copula.

A single-tier analysis is one where the copula verb is not required or not permitted, and the copula predicate is taken to select for a subject. This is illustrated in the f-structure in (42) for the translation of a Japanese sentence meaning 'the book is red', from Dalrymple et al. (2004: 191). The copula verb, if present, may contribute tense, as seen in Japanese (Dalrymple et al. 2004).
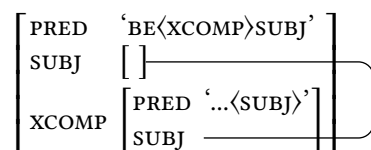
(42) Single-tier analysis (Dalrymple et al. 2004: 191)

$$\begin{bmatrix} \text{PRED} & \text{`RED}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & [\text{PRED `BOOK'}] \end{bmatrix}$$
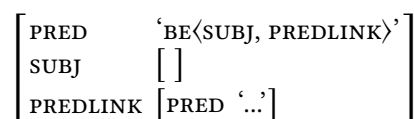
Jenny Graver

In a double-tier analysis, the copula provides the main predicate of the clause, and selects for either an open XCOMP function, or the closed PREDLINK function ("closed" meaning here that PREDLINK does not allow functional control).[13]

(43)   Double-tier analyses (Dalrymple et al. 2004: 189)

    a. Open complement

$$\begin{bmatrix} \text{PRED} & \text{`BE}\langle\text{XCOMP}\rangle\text{SUBJ'} \\ \text{SUBJ} & [\ ] \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`...}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \end{bmatrix} \end{bmatrix}$$

    b. Closed complement

$$\begin{bmatrix} \text{PRED} & \text{`BE}\langle\text{SUBJ, PREDLINK}\rangle\text{'} \\ \text{SUBJ} & [\ ] \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{`...'} \end{bmatrix} \end{bmatrix}$$

Sulger's (2009) argument in favour of a double-tier, closed complement analysis of the Irish copula is twofold, and has to do with the presence or absence of the copula, and the presence or absence of agreement between the copula predicate and subject.

Sulger (2009: 570) refers to the discussion between Dalrymple et al. (2004) and Attia (2008) on what to take away from the presence or absence of the copula. Ó Siadhail (1989: 244) formulates the general rule for the Irish copula as follows: "(...) the copula may not normally be deleted when marked for mood, tense, negation, interrogation or when embedded in a sentence." This is illustrated in (44):

(44)   Irish (Ó Siadhail 1989: 244)

    a. Múinteoir é   an fear sin.
        teacher   AGR.3SG.M DEF man that
        'That man is a teacher.'

    b. Ba   mhúinteoir é.
        COP.PST teacher   he
        'He was a teacher.'

---

[13]A reviewer provided examples from Welsh where the copula occurs with an expletive subject said to be required by the complement. The examples appear to involve modal semantics. More work is needed on how this fact should be analysed in light of the above discussion on the different analyses of the syntax of the copula. Irish has periphrastic modal predicates with the copula, and for Irish my intuition would be that these would need to be treated separately from regular copula predication as discussed in this chapter. See Graver (2010: 86–94) for an overview of Irish modal verbs with references for further reading.

c. Deir   siad gur   duine deas é.
   say.PRS they that.COP person nice he
   'They say that he is a nice person.'

On the basis of these facts, Sulger (2009: 570) argues that the absence of the copula, in the contexts where it may be dropped, is a matter of stylistic variation, and that the presence or absence of the copula does not lead to semantic differences.

Sulger suggests that his argument runs counter to Dalrymple et al. (2004). Dalrymple et al. (2004: 190–191) show how the Japanese copula may be dropped with adjectival predicates but is mandatory with nominal predicates. They argue on the basis of syntactic criteria that the category of the predicate may affect whether it can license a subject and propose a single-tier analysis for Japanese copula sentences with adjectival predicates whether or not the copula is present. For Japanese copula clauses with nominal predicates, they suggest a double-tier analysis of some kind. Sulger on the other hand argues on the basis of Attia (2008) that the predication is the same independent of the presence or absence of the copula, and for this reason that a unified analysis is desirable.

For a language like Russian, where the occurrence of copula is governed by tense, Dalrymple et al. (2004: 191–193) suggest that a unified analysis is desirable in this case, independent of the presence or absence of the copula. The point in this case is that there should not be any evidence of syntactic or semantic differences between clauses with the copula and clauses without. This is likely the case in Irish. Such a unified analysis would take two forms, either a single-tier analysis like Japanese, with the copula contributing features of tense, or a double-tier analysis with the copula as the main PRED of the clause selecting for either an XCOMP or a PREDLINK.

Sulger goes on to note that agreement between the copula predicate and the subject has been given by Dalrymple et al. (2004) as an argument in favour of an XCOMP analysis, because they view "agreement as a strong indication for a control relation between the subject and the predicate" (Sulger 2009: 566). There is no agreement between the copula predicate and the subject in Irish (see Mac Eoin 2002: 115 on the use of adjective predicates with the copula; for nouns compare (45) with (44) above).[14] Consequently, Sulger (2009: 567) argues, agreement is not an argument in favour of an XCOMP analysis in Irish.

(45)   Irish (Ó Siadhail 1989: 224)

---

[14]The pronominal element glossed AGR in some of these examples is inserted to agree with the subject, and cannot be taken to involve agreement between the subject and the predicate. See Carnie (1997: 61) and Ó Siadhail (1989: 224).

Jenny Graver

> Is múinteoir í   Cáit
> COP teacher   AGR.3SG.F Cáit
> 'Cáit is a teacher.'

Not all the linguistic literature on the Celtic languages in other theories agrees with this analysis. For example, Carnie & Harley (1994) provide a Principles and Parameters analysis of certain facts of the two Irish 'be' verbs where they view the copula as a complementiser particle providing features of aspect and tense (see Doherty 1996: 9–10 for arguments in favour of such an analysis based on how the copula behaves in sentences with interrogation, negation and subordination particles).[15] In LFG terms, this might imply a single-tier analysis.

There are hints in the philological studies and grammars of Old Irish that a single-tier analysis might be appropriate for the older stages of Irish and Scottish-Gaelic, and perhaps also for earlier stages of Welsh. For example, Ahlqvist (1971-72: 271) calls the copula a "verb-making particle", and Thurneysen (1998: 24–25) and McCone (1996: 211) discuss the similarities between the Old Irish copula and proclitic elements like pre-verbs and articles. Fife, in his introduction to the edited volume *The Celtic Languages*, writes as follows (Fife 2002: 20): "[f]ormerly, in both Irish and Welsh, the copula and its predicate formed a constituent, with the subject moved rightward to the end of the clause." Another point to note is the fact that Old Irish showed agreement between the subject and an adjective predicate in copula clauses. There is in other words much more Irish material to study when it comes to copula clauses.

5.3 Stage level and individual level predication

Sulger (2011) provides an analysis of copula constructions that express possession in Irish and Hindi/Urdu. For Irish he shows how the copula and the substantive verb behave in terms of the distinction between stage level and individual level predication. He argues that this contrast is expressed through lexical information. Specifically, he suggests that the substantive verb may supply a situation argument (based on Kratzer 1995) when it expresses stage level predication. The situation argument serves to embed the property expressed by the predication in some situation.

Sulger (2011: 19–20) again assumes a syntactic analysis using the PREDLINK function of the Irish copula, as mentioned in the previous section. For reasons of space he does not provide any examples of f-structures or lexical entries for his Irish data.

---

[15]See Asudeh (2002) for a general analysis of Irish pre-verbal particles.

In the following I will use Sulger's (2011) data as a starting point for illustrating how the copula and the substantive verb behave in terms of the stage and individual level distinction.

Sulger (2011: 12) notes that the linguistic literature on Irish generally assumes that the copula expresses individual level predication and the substantive verb stage level predication (see e.g. Doherty 1996: 40). "Stage level" in this sense refers to properties that hold of an individual at some stage of their lives, whereas "individual level" refers to properties that holds of an individual at all stages. The contrast between the copula and individual level predication, and the substantive verb and stage level predication, is nicely illustrated by MacEoin:

(46) Irish (Mac Eoin 2002: 136)

    a. Is   dochtúir mise
       cop doctor   I.emph
       'I am a doctor.'

    b. Tá   mise i mo dhocthúir
       be.prs I.emph in my doctor
       'I am a doctor.'

(46)a is a sentence with the copula verb is. (46)b on the other hand contains the substantive verb tá, with a subject 'I' and a prepositional phrase with the preposition 'in' together with a possessive particle 'my'. Mac Eoin (2002: 136) describes the differences between these examples as follows: "[...] *Is dochtúir mise* [with the copula] is an absolute statement of what I am, whereas *Tá mise i mo dhochtúir* [with the substantive verb] merely states the role in which I appear."

He goes on to contrast the above examples with the following:

(47) Irish (Mac Eoin 2002: 137)

    a. Is   gunna é   seo
       cop gun   this
       'This is a gun.'

    b. *Tá   sé seo ina ghunna
       be.prs this in.its gun
       Intended: 'This is a gun.'

In the latter example above, the construction with the substantive verb + 'in' + possessive particle cannot be used with 'gun' as the subject, since 'being a gun' is an absolute property of the thing referred to.

Jenny Graver

Sulger (2011) tests the claim that the Irish copula expresses individual level predication on a certain type of copula sentence in comparison with his Hindi data. Sulger terms the construction in question the possessive copula construction, where the copula is followed by a prepositional phrase with the preposition *le* 'with' expressing the possessor and a noun expressing the possessee. Sulger then applies some of the well-known tests for stage or individual level predication (Sulger 2011: 7 and references therein). For example, stage level predicates are assumed to allow temporal adverbs while individual level predicates do not. This is illustrated in (48)a, where, according to Sulger, the copula sentence is judged as questionable by native speakers with the addition of the adverb *inniu* 'today'.

Another test described by Sulger is to change the tense of a sentence, which is thought to result in a change in the perceived lifetime of the individual(s) involved in an individual level predication, but not in a stage level predication. This is illustrated for Irish in (48)b, which now implies that either Pádraig or the car does not exist anymore.

(48) Irish (Sulger 2011: 12, 14)

    a. Is   le   Pádraig an   carr nua (?inniu).
        COP.PRS with Pádraig DEF car   new today
        'Pádraig has the new car today.'

    b. Ba   le   Pádraig an   carr nua.
        COP.PST with Pádraig DEF car   new
        'Pádraig had the new car.'

For the substantive verb on the other hand, Sulger (2011: 15) points out, referencing Doherty (1996), that while a change in tense in examples similar to those in (48) results in the subject being perceived as dead when the copula is used, with the substantive verb the subject might have changed profession.

Sulger (2011: 12–14) goes on to show that while the copula is restricted to individual level predication, the substantive verb may in fact express both stage and individual level predication. For example, in the following example with the substantive verb, the reading is ambiguous between ownership (individual level) and temporary possession (stage level):

(49) Irish (Sulger 2011: 12)
     Tá   an carr nua ag Pádraig
     be.PRS DEF car new at Pádraig
     'Patrick has the new car' (he may or may not own it)

6 Conclusion

I hope to have shown that the work in LFG on the Celtic languages, while not very substantial, has contributed in various ways to both the theory of LFG and to our understanding of the languages themselves. For example, the question of whether there is a VP in a VSO language like Welsh has been drawn into the discussion of endocentricity and extended heads in LFG (Section 2.2), and the autonomous verb form in Irish has been analysed in the context of general, cross-linguistically applicable categories describing relationships between thematic roles and syntactic functions (Section 3.2).

At the same time, there is a lot of material in the Celtic languages remaining to be studied for the interested researcher. Does it take some extra dedication from the non-native speaker researcher especially, given the challenges of working on minority languages, the low number of native speakers and comparative lack of teaching materials? Yes. But I would still argue that it is very much worth it.

Acknowledgments

Abbreviations

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

AUT	autonomous verb form	IMPERS	impersonal
CONTR	contrastive	PRED	predicative particle
EMPH	emphatic		

References

Ahlqvist, Anders. 1971-72. Some aspects of the copula in Irish. *Éigse* 14. 269–274.

Anderson, Stephen R. & Sandra Chung. 1977. On grammatical relations and clause structure in verb-initial languages. In Peter Cole & Jerrold M. Sadock (eds.), *Grammatical relations* (Syntax and Semantics 8), 1–25. New York: Academic Press.

Jenny Graver

Andrews, Avery D. 1990. Unification and morphological blocking. *Natural Language & Linguistic Theory* 8(4). 507–557. DOI: 10.1007/bf00133692.

Arman, Laura. 2015. *The Welsh impersonal construction.* Manchester: University of Manchester. (Doctoral dissertation).

Asudeh, Ash. 2002. The syntax of preverbal particles and adjunction in Irish. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '02 conference.* Stanford: CSLI Publications.

Asudeh, Ash. 2012. *The logic of pronominal resumption* (Oxford Studies in Theoretical Linguistics). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199206421.001.0001.

Attia, Mohammed. 2008. A unified analysis of copula constructions in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '08 conference*, 89–108. Stanford: CSLI Publications.

Ball, Martin J. & Nicole Müller. 1992. *Mutation in Welsh.* London/New York: Routledge. DOI: 10.4324/9780203192764.

Belyaev, Oleg. Forthcoming. Core concepts of LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar.* Berlin: Language Science Press.

Blevins, James. 2003. Passives and impersonals. *Journal of Linguistics* 39. 473–520. DOI: 10.1016/b0-08-044854-2/00292-3.

Borsley, Robert D. 2006. On the nature of Welsh VSO clauses. *Lingua* 116(4). 462–490. DOI: 10.1016/j.lingua.2005.02.004.

Borsley, Robert D. 2013. On the nature of Welsh unbounded dependencies. *Lingua* 133. 1–29. DOI: 10.1016/j.lingua.2013.03.005.

Borsley, Robert D. 2019. The complexities of the Welsh copula. In Stefan Müller & Petya Osanova (eds.), *Proceedings of the 26th international conference on Head-Driven Phrase Structure Grammar*, 5–25. Stanford: CSLI Publications.

Borsley, Robert D. 2020. On a family of Welsh constructions. In Stefan Müller & Anke Holler (eds.), *Proceedings of the 27th international conference on Head-Driven Phrase Structure Grammar*, 26–46. Stanford: CSLI Publications. DOI: 10.21248/hpsg.2020.2.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh.* Cambridge, UK: Cambridge University Press. DOI: 10.1017/cbo9780511486227.

Bresnan, Joan. 2001. *Lexical-Functional Syntax.* Oxford: Blackwell.

Broadwell, George Aaron. 2005. It ain't necessarily S(V)O: Two kinds of VSO languages. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '05 conference*, 73–83. Stanford: CSLI Publications.

Carnie, Andrew. 1997. Two types of non-verbal predication in Modern Irish. *Canadian Journal of Linguistics* 42(1–2). 57–73. DOI: 10.1017/s0008413100016820.

Carnie, Andrew. 2005. Flat structure, phrasal variability and VSO. *Journal of Celtic Linguistics* 9. 13–31.

Carnie, Andrew & Eithne Guilfoyle. 2000. Introduction. In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*, 3–11. Oxford: Oxford University Press. DOI: 10.1075/la.62.03car.

Carnie, Andrew & Heidi Harley. 1994. Nominally phrasal copular constructions. In Vida Samiian & Jeanette Schaeffer (eds.), *Proceedings of the 24th Western Conference on Linguistics*, 30–43. Fresno: California State University.

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.

Doherty, Cathal. 1996. Clausal structure and the Modern Irish copula. *Natural Language & Linguistic Theory* 14. 1–46. DOI: 10.1007/bf00133402.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the world*. 22nd edn. Dallas: SIL International.

Fife, James. 2002. Introduction. In Martin J. Ball & James Fife (eds.), *The Celtic languages*. London/New York: Routledge. First paperback edition.

Findlay, Jamie Y. & Anna Kibort. Forthcoming. Argument structure and mapping theory. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Graver, Jenny. 2010. *Passives and impersonals: A diachronic study of the Irish autonomous verb in the framework of Lexical Functional Grammar*. Oslo: University of Oslo. (Doctoral dissertation).

Graver, Jenny. 2011. The syntax and development of the Old Irish autonomous verb. In Andrew Carnie (ed.), *Formal approaches to Celtic linguistics*, 41–63. Newcastle upon Tyne: Cambridge Scholars Publishing.

Graver, Jenny. Forthcoming. LFG and Celtic languages. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Harlow, Steve. 1989. The syntax of Welsh soft mutation. *Natural Language & Linguistic Theory* 7. 289–316. DOI: 10.1007/bf00208099.

Haug, Dag. Forthcoming. Agreement. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Jones, Bob Morris. 2002. The 'genitive of respect' revisited. Abstract of paper presented at the Ninth Welsh Syntax Seminar, Plas Gregynog. https://users.aber.ac.uk/bmj/Wss/Sae/wss9.html.

Kaplan, Ronald M. & Annie Zaenen. Forthcoming. Long-distance dependencies. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Jenny Graver

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.

Kennard, Holly J. 2014. The persistence of verb second in negative utterances in Breton. *Journal of Historical Linguistics* 4(1). 1–39. DOI: 10.1075/jhl.4.1.01ken.

Kennard, Holly J. 2019. Morphosyntactic and morphophonological variation in Breton: A cross-generational perspective. *Journal of French Language Studies* 29. 235–263. DOI: 10.1017/s0959269519000115.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '07 conference*, 250–270. Stanford: CSLI Publications.

Kibort, Anna. 2014. Mapping out a construction inventory with (Lexical) Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '14 conference*, 262–282. Stanford: CSLI Publications.

King, Tracy Holloway & Mary Dalrymple. 2004. Determiner agreement and noun conjunction. *Journal of Linguistics* 4(1). 69–104. DOI: 10.1017/s0022226703002330.

Kratzer, Angelika. 1995. Stage-level and individual-level predicates. In Gregory N. Carlson & Francis Jeffry Pelletier (eds.), *The generic book*, 125–175. Chicago: University of Chicago Press.

Mac Eoin, Gearóid. 2002. Irish. In Martin J. Ball & James Fife (eds.), *The Celtic languages*. London/New York: Routledge. First paperback edition.

McCloskey, James. 1979. *Transformational syntax and model theoretic semantics: A case-study in Modern Irish*. Dordrecht: Reidel.

McCloskey, James. 1983. A VP in a VSO language? In Gerald Gazdar, Ewan Klein & Geoffrey K. Pullum (eds.), *Order, concord and constituency*, 9–55. Dordrecht: Foris Publications. DOI: 10.1515/9783110876734-003.

McCloskey, James. 1986. Inflection and conjunction in Modern Irish. *Natural Language & Linguistic Theory* 4. 245–281. DOI: 10.1007/bf00133846.

McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and explanation in the minimalist program*, 184–226. Malden: Blackwell. DOI: 10.1002/9780470755662.ch9.

McCloskey, James. 2003. Working on Irish. *GLOT* 73. 63–72. https://people.ucsc.edu/~mcclosk/PDF/glot.pdf.

McCloskey, James. 2007. The grammar of autonomy in Irish. *Natural Language & Linguistic Theory* 25. 825–857. DOI: 10.1007/s11049-007-9028-7.

McCloskey, James & Kenneth Hale. 1984. On the syntax of person-number inflection in Modern Irish. *Natural Language & Linguistic Theory* 1. 487–533.

McCone, Kim. 1996. An tSéan-Ghaeilge agus a réamhstair. In Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach (eds.), *Stair na Gaeilge. In ómós do Pádraig ó Fiannachta*. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.

Mittendorf, Ingo & Louisa Sadler. 2005. Numerals, nouns and number in Welsh NPs. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '05 conference*, 294–312. Stanford: CSLI Publications.

Mittendorf, Ingo & Louisa Sadler. 2006. A treatment of Welsh initial mutation. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '06 conference*, 343–364. Stanford: CSLI Publications.

Mittendorf, Ingo & Louisa Sadler. 2008. NP would like to meet GF: A Welsh adjectival construction. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '08 conference*, 373–393. Stanford: CSLI Publications.

Ó Siadhail, Mícheál. 1982. Cardinal numbers in Modern Irish. *Ériu* 33. 99–107.

Ó Siadhail, Mícheál. 1989. *Modern Irish: Grammatical structure and dialectal variation*. Cambridge/New York/Melbourne: Cambridge University Press.

Sadler, Louisa. 1997. Clitics and the structure-function mapping. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '97 conference*. Stanford: CSLI Publications.

Sadler, Louisa. 1998. Welsh NPs without head movement. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '98 conference*. Stanford: CSLI Publications.

Sadler, Louisa. 1999. Non-distributive features in Welsh coordination. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '99 conference*. Stanford: CSLI Publications.

Sadler, Louisa. 2006. Function spreading in coordinate structures. *Lingua* 116(11). 1777–1806. DOI: 10.1016/j.lingua.2004.09.005.

Schafer, Robin. 1995. Negation and verb second in Breton. *Natural Language & Linguistic Theory* 13. 135–172. DOI: 10.1007/bf00992780.

Schmidt, Karl Horst. 2002. Insular Celtic: P and Q Celtic. In Martin J. Ball & James Fife (eds.), *The Celtic languages*. London/New York: Routledge. First paperback edition.

Sproat, Richard. 1985. Welsh syntax and VSO structure. *Natural Language & Linguistic Theory* 3. 173–216. DOI: 10.1007/bf00133840.

Stenson, Nancy. 1989. Irish autonomous impersonals. *Natural Language & Linguistic Theory* 7. 379–406. DOI: 10.1007/bf00208102.

Stephens, Janig. 2002. Breton. In Martin J. Ball & James Fife (eds.), *The Celtic languages*. London/New York: Routledge. First paperback edition.

Jenny Graver

Sulger, Sebastian. 2009. Irish clefting and information-structure. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '09 conference*, 562–582. Stanford: CSLI Publications.

Sulger, Sebastian. 2010. Analytic and synthetic verb forms in Irish – an agreement-based implementation in LFG. In Manfred Pinkal, Ines Rehbein, Sabine Schulte im Walde & Angelika Storrer (eds.), *Semantic approaches in natural language processing: Proceedings of the Conference on Natural Language Processing 2010*, 169–174. Saarbrücken: Saarland University Press.

Sulger, Sebastian. 2011. A parallel analysis of two have-type copular constructions in two have-less Indo-European languages. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG '11 conference*, 503–523. Stanford: CSLI Publications.

Tallerman, Maggie. 1990. Relativization strategies: NP accessibility in Welsh. *Journal of Linguistics* 26(2). 291–314. DOI: 10.1017/s0022226700014687.

Tallerman, Maggie. 1998. Word order in Celtic. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 21–46. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110812206.21.

Tallerman, Maggie. 2006. The syntax of Welsh "direct object mutation" revisited. *Lingua* 116(11). 1750–1776. DOI: 10.1016/j.lingua.2004.10.004.

The Christian Brothers. 2002. *New Irish grammar*. Dublin: C. J. Fallon. Originally published in 1962.

Thurneysen, Rudolf. 1998. *A grammar of Old Irish*. Dublin Institute for Advanced Studies.

Timm, Leonora A. 1989. Word order in 20th century Breton. *Natural Language & Linguistic Theory* 7. 361–378. DOI: 10.1007/bf00208101.

Watkins, T. Arwyn. 2002. Welsh. In Martin J. Ball & James Fife (eds.), *The Celtic languages*. London/New York: Routledge. First paperback edition.

Wechsler, Stephen & Larisa Zlatić. 2000. A theory of agreement and its application to Serbo-Croatian. *Language* 76(4). 799–832. DOI: 10.2307/417200.

Handbook of Lexical Functional Grammar

Set blurb on back with \BackBody{my blurb}